Indonesian Personal Income Tax Microsimulation:

Tax Base Construction, Revenue, Distribution

and Compliance Analysis

(Case Study of Income Tax Reform 2008)

By:

Bimo Wijayanto

Supervisory Panel:

Professor Rob Tanton (Chair)

Professor Alan Duncan

Dr Yogi Vidyattama

Faculty of Business and Government

University of Canberra

2015
ABSTRACT

Following the 1980’s oil price plunge, Indonesian economic growth fell sharply as a consequence of its heavy reliance on oil and gas exports. This drove the long overdue need to reform the tax law to replace an old and outdated colonial tax. This was the beginning of radical move to a self-assessment and simplification of the tax system to increase future revenue streams and to repay an increasing foreign debt. Since 1983, the tax policy has undergone ongoing reforms to improve effectiveness, efficiency, neutrality, vertical and horizontal equity with the final aim to encourage voluntary compliance to increase government revenue. Somehow, after years of reform, the actual tax performance is still far from what was initially expected. Among the ASEAN countries, Indonesia currently experiences the lowest tax ratio to GDP. Due to this low tax ratio, the total contribution of income tax (Indonesia’s main tax revenue) is smaller than other ASEAN countries.

Indonesian income tax law amendment typically is a long process with intensive discussion between the parliament and the government. It was observed that the government mainly used macro-based analysis. This analysis will be more cost effective and suitable when the primary concern is the aggregate-impact. However, several study have endorsed microsimulation model as the policy tool to address issue related to the need for detail impact of the policy scenarios. This model captures the interaction between the individual in the tax system and the assessment of the relational impact with other welfare program. This has led to the initial motivation of this thesis, the need for the development of a sound microsimulation model for the purpose of personal income tax (PIT) policy analysis.
A static microsimulation model is built for the current structure of the Indonesian PIT. We specifically looked at four major research areas - the best data source for constructing the base file of the Indonesian PIT (potential number of taxpayers and income base of the Indonesian PIT); the construction of PIT microsimulator and its sensitivity test; the impact of PIT reform on potential revenue, the distribution of the tax burden and the potential tax gap; the compliance issue and how it plays a part in the revenue accumulation.

This thesis builds a better database and elaborates a method for Indonesian PIT base file construction. We apply recommendations from the evaluation of several developed, transitional and developing countries’ microsimulation models. We conduct a preliminary attempt to improve the PIT base file by using statistical matching of different datasets. This attempt overcomes the problems and limitations found in available survey and administrative data. Three sets of the Indonesian micro data were used -The Socio Economic Survey, Labour Force Survey and tax administrative data. The first two database were combined before being imputed with the latter. This microsimulation is a powerful tool for the Indonesian government to design its tax policy. This empirical analysis aims to fill the gap in the knowledge of microsimulation in Indonesia.
ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to Allah SWT, God the Almighty whom has given me all his blessings and mercy during my life and especially while I am conducting this study at UC.

My sincere thanks to my beloved parents; Mama Aniek Andjarwati and Papa Samidjo, my dearest wife, Palupi Retnaning Kusumastuti and our children, the love of our life: Aisyapadma Btari Pramatya Wijayanto, Asyampadma Bharata Pandu Wijayanto and Alanzapadma Bestari Pramastri Wijayanto, my sisters; Ratna Damayanti and Niken Tri Wulandari. Its all the quintessence of your endless love, patience, prayers and your supports to my journey. I would have never made it without you. This thesis is dedicated to you.

My deepest thanks to my supervisors: Dr Yogi Vidyattama, Professor Robert Tanton (my chair) and Professor Alan Duncan (especially during his time at NATSEM from 2010-2013), for all of their wisdom, encouragement, and for their supportive and constructive supervision. I am truly grateful to work with them.

I would also like to acknowledge my previous supervisor: Professor Ann Harding; Dr Quoc Ngu Vu who gave me their preliminary insight for the initial stage of my research. Thanks to all of my confirmation seminar reviewer: Prof. Phil Lewis, Prof. Deborah Blackman, Prof. Anne Daly for their valuable comments during the different stage of this study. Thanks to Dr. Denis Whitfield, for proofreading my thesis.

I am indebted to Thalyta Yuwono, Rubino Sugana and Stephen Clark who introduced me to microsimulation. Thank you so much to Indonesian Directorate General of Taxes and Bureau of Statistics, Nopi Kurniawan, Andri Parwito, Ahmad Riswan for providing the data. Thanks
for the great support from all friends and colleagues in the DJP, Kemenkeu and the BPS of the Republic of Indonesia.

I acknowledge the Department of Foreign Affairs Australia (previously AusAid) through their Australian Leaderships Award and Hadi Soesastro Prize-Australia Awards for such generous financial support throughout my study in NATSEM at the University of Canberra and last tiny piece of my study at the Duke University. I would also like to acknowledge my scholarships referee: Dr Bambang Riyanto LS, Dr Sumihar Petrus Tambunan, Mas Phillip Ironfield and Campher Serfontein. Special thanks to UC AusAid SCO: Rozana Muir for all helpful advice and warm support throughout my study period. Thanks also to UC International Office for forwarding my application for HSP-AA.

I would like to extend my thank to my brothers and sisters who shared their great and productive time with me, during my study in Canberra. Our brotherhood will be forever: IH 4-ers: Toki, Liz, Juned, Kukun, Intan, Farid, Hata for all the times we shared in my last 1.5 years at the cozy International House 4 at Cooinda Hut.

UCISS, UCKum, Mofilink, Kagama Canberra and all Canberrans from 2010-2014: thanks so much for the great friendships during the years.

Finally, I also like to thank my parents-in-law, brothers and sisters in law in Singapore, Jakarta, Wonogiri and Yogyakarta for all their prayers and supports.
Contents

FORM B .................................................................................................................................................. I

CERTIFICATE OF AUTHORSHIP OF THESIS ...................................................................................... I

ABSTRACT ............................................................................................................................................... III

ACKNOWLEDGMENTS ............................................................................................................................ V

LIST OF TABLES .................................................................................................................................. XI

CHAPTER I. INTRODUCTION ................................................................................................................. 1
  1.1 BACKGROUND ................................................................................................................................. 1
  1.2 MOTIVATION/AIM OF THE STUDY ................................................................................................. 5
    1.2.1 General Research Questions ...................................................................................................... 8
    1.2.2 Specific Research Questions ...................................................................................................... 8
    1.2.3 Contributions ............................................................................................................................. 9
  1.3 REVIEW OF LITERATURE AND DATA .......................................................................................... 10
    1.3.1 Microsimulation Modelling ....................................................................................................... 10
    1.3.2 Application in Different Countries .............................................................................................. 15
  1.4 THE AVAILABLE DATA SOURCE ................................................................................................ 21
  1.5 MODEL CONSTRUCTION STRATEGY .............................................................................................. 22
  1.6 THESIS OUTLINE .......................................................................................................................... 24

Based on its intended contribution, the thesis is organized as follows: ............................................. 24

I. Introduction ....................................................................................................................................... 24
II. Income in Indonesian Micro Data .................................................................................................... 25
III. Developing Indonesian Labour Income Tax Microsimulation ....................................................... 25
IV. The Revenue and Distribution Impact Analysis of Indonesian Personal Income Tax Reform in 2008 25
V. Conclusion ....................................................................................................................................... 26

CHAPTER II. INCOME IN INDONESIAN MICRO DATA ......................................................................... 27

2.1 INTRODUCTION ............................................................................................................................ 27
2.1.1 The need of income for modelling tax microsimulation .......................................................... 29

2.2 Indonesian income micro data .................................................................................................. 31

2.2.1 SUSENAS .......................................................................................................................... 32

History of the survey ..................................................................................................................... 32
Survey methodology ...................................................................................................................... 33
Questionnaire and variables ......................................................................................................... 34
Coverage .................................................................................................................................. 35

2.2.2 SAKERNAS ....................................................................................................................... 36

History of the survey ..................................................................................................................... 36
Survey methodology ...................................................................................................................... 36
Questionnaire and variables ......................................................................................................... 37
Coverage .................................................................................................................................. 39

2.2.3 Administrative data/income tax return .................................................................................. 40

Variables .................................................................................................................................. 41
Coverage .................................................................................................................................. 43

2.2.4 Data comparison (Strengths and weaknesses from each micro data) ...................................... 43

SUSENAS .................................................................................................................................. 43
SAKERNAS .................................................................................................................................. 47
Administrative data/income tax return ......................................................................................... 49

2.3 Filling the gap in the income survey data ................................................................................ 50

2.3.1 Harmonisation and reconciliation of sources ..................................................................... 53

Target population ........................................................................................................................ 54
Comparison of concept and distribution of common variables .................................................. 57
Demographic: highest educational level attained ........................................................................ 61
Demographic: age group and sex ................................................................................................ 62
Demographic: marital status ........................................................................................................ 64
Demographic: number of household members ........................................................................... 64
Employment: active working status ............................................................................................ 65
Employment: main occupation sectors ....................................................................................... 67
Wage concept and reference period ............................................................................................ 68

2.3.2 Income alignment and grossing up ....................................................................................... 69
CHAPTER III. DEVELOPING INDONESIAN LABOUR INCOME TAX MICROSIMULATION

3.1 INTRODUCTION

3.1.1 Tax Microsimulation in Indonesia

3.1.2 Utilisation

3.2 PERSONAL INCOME TAX (PIT) IN INDONESIA

3.2.1 Definition and Coverage

3.2.2 History of IT-Based Tax Records in Indonesia

3.3 METHODOLOGY

3.3.1 Scope and Ruling

3.3.2 Variables and Determinants

3.3.3 Adjustment and Formula for PIT Simulator

3.4 APPLICATION OF TAX MICROSIMULATION IN THE INDONESIAN INCOME TAX POLICY

3.4.1 Initial Tax Microsimulation on un-modified Survey Data
3.4.2 Initial Tax Microsimulation on modified Survey Data .................................................. 123
Susenas Imputed with Sakernas Labour Income.................................................................. 123
Susenas Imputed with Sakernas Labour Income and PIT Returns ..................................... 125
3.4.3 What we learned from tax microsimulation ............................................................... 127
Comparison of the modified and unmodified databases..................................................... 127
The Contribution of Higher Income Earners from the Tax Administrative Data in the Modified Database........... 131
3.5 The Sensitivity of the Simulator to Taxpayers’ Dependents........................................... 131
3.6 Application for Future Policy Simulation ..................................................................... 135
3.6.1 The Benefit from Applying Different Maximum Dependent for Tax Relief Component .......... 136
3.6.2 The Benefit from Different Taxable Income Threshold............................................... 138
3.7 Conclusion.................................................................................................................... 139

CHAPTER IV. REVENUE AND DISTRIBUTION IMPACT ANALYSIS OF INDONESIAN PERSONAL INCOME TAX

REFORM IN 2008..................................................................................................................... 145

4.1 Introduction..................................................................................................................... 145
4.1.1 Aim of tax in Indonesia.............................................................................................. 147
4.1.2 The detail of recent reform......................................................................................... 152
4.1.3 The performance of PIT in Indonesia........................................................................ 156
4.2 Empirical Framework.................................................................................................... 160
4.3 Results and Discussions............................................................................................... 162
4.3.1 The impact on the tax basis....................................................................................... 162
4.3.2 The Impact on Revenue ........................................................................................... 164
4.3.3 The Impact on the Tax Burden Distribution.............................................................. 168
4.4 Tax Potential Versus Tax Compliance ........................................................................ 170
4.5 Conclusion..................................................................................................................... 173

CHAPTER V. CONCLUSIONS AND FINDINGS...................................................................... 175

5.1 Introduction..................................................................................................................... 175
5.2 Findings........................................................................................................................ 177
5.2.1 Building Indonesia income base file ....................................................................... 177
5.2.2 Building Indonesian PIT microsimulation ................................................................. 180
5.2.3 Revenue and distribution impact (PIT reform in 2008) .................................................. 182
5.3 IMPLICATIONS FOR POLICY ......................................................................................... 183
5.4 LIMITATIONS OF THE CURRENT STUDY ..................................................................... 184
  5.4.1 Problems of the Indonesian Income Data ................................................................. 184
  5.4.2 The issue of compliance ............................................................................................ 186
  5.4.3 Static microsimulation to answer the research questions in this study ...................... 186

REFERENCES .......................................................................................................................... 189

List of Tables

Table 1. Comparison of weighted frequency for individual reference population by province between Sakernas & Susenas 2008 .................................................................................................................................................. 56
Table 2. Difference and Harmonisation of Variables’ Conceptual Differences and their Sub-Coding (Sakernas & Susenas 2008) ........................................................................................................................................ 60
Table 3. Comparison of weighted frequency for individual reference population after harmonisation of highest educational level variable (Sakernas & Susenas 2008) ........................................................................ 61
Table 4. Comparison of weighted frequency for individual reference population by age group and sex variable (Sakernas & Susenas 2008) ............................................................................................................. 63
Table 5. Comparison of weighted frequency for individual reference population by marital status variable (Sakernas & Susenas 2008) ............................................................................................................ 64
Table 6. Comparison of weighted frequency for individual reference population by number of household member (Sakernas & Susenas 2008) ........................................................................................................... 65
Table 7. Comparison of weighted frequency for individual reference population by active working status (Sakernas & Susenas 2008) ................................................................................................................. 66
Table 8. Comparison of weighted frequency for individual reference population by main occupation sectors (Sakernas & Susenas 2008) .............................................................................................................. 67
Table 9. Difference between national publications on average labour wages per province vs weighted average labour wages per province (SAKERNAS 2008 Micro Data) ........................................................................... 69
Table 10. Correlation between imputing variables and income variables for Sakernas 2008 .78
Table 11. Summary Statistics: Sakernas raw data compared to Susenas imputed using Modified cold-Deck and Hot-deck Method .................................................................................................................. 84
Table 12. Summary of Matching Results by Decile of Income .................................................. 85
Table 13. Summary of Matching Results by Province ............................................................... 87
Table 14. Summary of Matching Results by Highest Education Level ...................................... 89
Table 15. Summary of Matching Results by Main Occupation Sectors ..................................... 90
Table 17. Article and Object of Personal Income Tax under Law no 36(2008) ....................... 106
Table 18. Determinants from PIT Return and SAKERNAS variables’ mapping .................. 114
Table 19. Initial simulation of the un-modified Raw Sakernas 2008 ........................................ 120
Table 20. Initial simulation of the Susenas imputed with Sakernas Labour Income 2008 ................................................................. 124
Table 21. Initial simulation of the modified Susenas imputed with Sakernas and Tax Return 2008 ................................................................. 126
Table 22. Average Effective Tax Rate between Sakernas Raw and the modified Susenas imputed with Sakernas 2008 ................................................................. 129
Table 23. Comparison of dependent distribution Sakernas 2008 and Susenas 2008 imputation ........................................................................ 133
Table 24. Different Dependent Simulation Susenas-Sakernas-PIT Imputed base file by decile ........................................................................ 134
Table 25. Revenue Impact and Distribution of Burden by decile from Different Dependent Susenas-Sakernas-PIT Imputed base file ........................................................................ 137
Table 26. Revenue and Distribution of Burden from Different Taxable Income Threshold Susenas-Sakernas-PIT Imputed base file by decile ........................................................................ 138
Table 27. Comparison: Share of Oil and Gas Revenue and National Tax Revenue to Domestic Revenue ........................................................................ 152
Table 28. PIT law changes since 1983 ........................................................................ 153
Table 29. Non-taxable income changes 1983-2008 ........................................................................ 155
Table 30. Number of Eligible Tax Payers based on Law no 17(2000)and Law no 36(2008) ........................................................................ 163
Table 31. Income Tax Revenue Simulation before and after Reform (Income Tax Denominated in Billion Rupiahs) ........................................................................ 165
Table 32. Distribution of Personal Income Tax Burden Before and After Reform ........................................................................ 168
Table 33. Tax Ratio and Personal Tax Payers Compliance ........................................................................ 171

Table of Figures

Figure 1. Share of tax revenue and detailed breakdown of income tax revenue (2009) .......... 3
Figure 2. SAKERNAS (2008) Grossing Up Process ........................................................................ 74
Figure 3. Step 1 to 3 of the matching process ........................................................................ 77
Figure 4. Step 4 to 5 of the matching process ........................................................................ 80
Figure 5. Plotting Income Salary Annual with the Population ........................................................................ 93
Figure 6. Plotting Kernel Density Estimate of the Annual Income from Salary/Wages ........ 94
Figure 7. Schematic of the PIT structure in Indonesia for the calculation of 2008 tax returns (Based on the Law no 17(2000) and Ministry of Finance Decree 37/PMK.03/2005) ........................................................................ 119
Figure 8. Tax as a percentage of GDP; ASEAN Countries (2008) ........................................................................ 157
Figure 9. Tax mix comparison with neighbouring countries (2007) ........................................................................ 158
CHAPTER I. INTRODUCTION

1.1 Background

In 1983, The Indonesian Government initiated the first fundamental tax reforms to replace the old colonial income tax law. Following the 1980’s oil price plunge, Indonesian economic growth fell sharply as a consequence of its heavy reliance on oil and gas exports. It drove the long overdue need to reform the tax law to increase future revenue stream and to repay an increasing foreign debt (Anwar, Aziz, Pangestu & Soesastro 1991; Gillis 1994; Heij 2000; Hill 2000; Resosudarmo & Kuncoro 2006). Tax had become a key element of support for the Indonesian government budget since 1983 (Ikhsan, Trialdi & Syahrial, 2005).

After the initial income tax reform of 1983, other reforms to improve tax law and administration reforms were continued sequentially; VAT (Value Added Taxes), stamp duty, land and building taxes. All were aimed at producing more neutral, simpler, and fairer tax laws. Tax has increasingly played an important role in the Indonesian government revenue. Currently it is the main revenue share of the Indonesian State Budget. During the last two decades, the tax revenue contribution increased significantly from only 39.5 percent in the 1980’s to an average of 70 percent from total domestic revenue since 2005. However, despite that performance, Indonesia had a relatively low tax ratio compared to the other ASEAN economies. Of nine ASEAN neighbours, only Cambodia and Myanmar had a lower tax to GDP ratio than Indonesia.

One type of tax that has become increasingly important in Indonesia is the income tax. There are two types of income tax – corporate and personal. Both corporate and personal income
tax is under the authority of the Indonesian central government. There is a specific treatment for PIT with regard to revenue sharing between the central and sub-national governments. The central government determines both the tax base and the tax rates for personal and corporate income tax through tax law amendments.

In Indonesia, income tax is classified under Non-Oil and Gas Income Tax revenue. Non-oil and gas income tax constituted the major proportion of total government revenue. In 2009, the non-oil and gas income tax revenue accounted for 41.2 percent of the total tax revenue. Of this, corporate income tax (CIT) revenue accounted for around 46.6 percent of the total non-oil and gas income tax while PIT article 21 (which is a labour income tax withheld by the employers) accounted for 18.8 percent of the total non-oil and gas income tax revenue. The other revenue of PIT (article 25/29 - from people with more than one source of income, individual business owner and professionals) has contributed only 1.2 percent of total non-oil and gas income tax revenue. This had led to a question about the potential revenue that could be generated from PIT. In a later chapter, a potential income tax estimation using a microsimulation model is developed. This model thoroughly examines the potential revenue and potential tax payer base from the available Indonesian personal income database.

To see how income tax, PIT in particular, contributes to the total tax, it is necessary to give a further breakdown (Figure 1). This figure shows how PIT contributes to the total tax revenue
Since its inception in 1983, Indonesia maintained simple personal income tax systems. There were no sophisticated means-tested allowances or benefits and no government transfer or subsidy attached to the personal tax systems. During 1983-2008, the main PIT law reforms were to change its progressivity. Changes were made to the income threshold, marginal tax rate and tax relief in the form of a non-taxable allowance.

The main issue facing the advancement of the PIT in Indonesia is the tax base. In Indonesia, personal income tax base is stipulated in the income tax law. All tax payers, whether they have one or more than one source of income, or whether they are salaried employees or professionals/traders/business-owner with business income and other income, must be registered in the system. They can either be registered through their employer or through self-registration. Each of them will get, a unique individual tax identification number. According to the Indonesian DGT administrative register, only 10 million individual tax payers were registered. This is despite a labour force survey by the Indonesian Bureau of Statistics (Sakernas BPS, 2009) showing that the Indonesian labour force was 113.74 million people.
with an active-working population of 104.49 million. A total of 73.12 million people, or 69.98 percent of the working population, worked for more than thirty five hours per week. This strongly suggests that the Indonesian tax authority still has a lot of scope to expand the Indonesian tax payer base. Ikhsan et al. (2005) argues that an expansion for a wider tax base can be implemented given that the current tax basis is very much concentrated in the highest income decile of tax payers.

Several researches have tried to analyse the possibility of increasing the income tax base especially in regards to compliance. Panjaitan (2005) applied pareto distribution analysis for three sources of income micro data (Susenas 1996; National Census 2000; Susenas 2002). Panjaitan’s results confirmed the Indonesian low tax compliance whereby only about 11 percent to 20 percent of potential revenue was covered by the Indonesian PIT system. Panjaitan further urged for the improvement of tax administration and tax payer education efforts to improve the compliance rate.

An International Monetary Fund (IMF) study by Brondolo, Silvani, LeBorgne & Bosch (2008) found that the relatively low tax ratio could imply a low tax burden in Indonesia’s tax system. They indicated that after twenty six years of reform Indonesia still had a narrow tax base and low taxpayer compliance. It seems that the problems are caused by several major weaknesses in Indonesian tax administration and its tax structure. This urgently needs government corrective action in order to mobilize adequate revenue to support Indonesia’s fiscal sustainability.
Another study, by the Organisation for Economic Co-operation and Development (OECD) (Gandullia, Iacobone & Thomas, 2012), found that Indonesia, together with India and South Africa, imposed a relatively low average and marginal PIT compared to the majority of OECD countries. The OECD suspected that only a fraction of high income workers in those countries face a substantial tax burden and it could explain their relatively low tax to GDP ratio compared to OECD countries. Ikhsan et al. (2005) argue that a high level of concentration of revenue streams from wealthy earners indicated a high level of vulnerability in the Indonesian income tax basis.

Following this background, I will give an overview of the aim of the study, its research questions, hypothesis and contributions. The third section is a brief literature review about the empirical framework used in this study; that is the use of microsimulation. This includes an overview of the microsimulation model, and examples of microsimulation studies across countries. Section four presents and discusses the available micro data, and the problem with the current data. Section five presents the proposed solution that can be expected from the work in this thesis followed by the plan of analysing the income tax reform. The last section is an outline of the thesis.

1.2 Motivation/Aim of the Study

In Indonesia, income tax law amendment is typically a long process. It requires intensive discussion between the Indonesian people’s representative agency (parliament) and the government. It took almost four years to finalise the last amendment. It started at the end of
2005 and continued until officially announced on 30 September 2008. Therefore, it is understandable there have been only three amendments to the personal income tax law since 1983.

It appears that macro analysis plays an important role in government discussions about tax. Each fiscal year, the Indonesian government and parliament establish a law for the national budget. In every related discussion, the government prepares a Nota Keuangan dan RAPBN (Financial Note and State Budget Revision Plan) of the Indonesian Ministry of Finance. The major analysis usually emphasizes the interaction between macro factors and government policy intervention to support the discussion of tax revenue monitoring and forecasting within the fiscal period. The impact of global and domestic economic performance is also included in the main agenda of discussion, together with the interaction of economic growth, crude-oil price, exchange rates and Indonesian oil and gas lifting production.

In some way, the aggregate level analysis is cost effective and suitable when the primary concern of policy makers is the tax policy impact on the macro-aggregate level. Learning from the panel evaluation on US microsimulation studies by Citro & Hanushek (1991), so far microsimulation models are the only alternative policy tools that can answer the need for detailed information on the impact of several policy scenarios on the public, the interaction of the individual components of the proposed system and also an assessment of the relational impact of the proposed policy with the various welfare programs. It is a more appropriate model for supporting evidence-based policy. It would satisfy the need for transparent and accountable tax policy reform. Those two are essential factors to satisfy the growing public awareness of the tax policy reform in Indonesia’s current democracy.
That leads to the initial motivation of this thesis; the need for a sound microsimulation model for income tax policy analysis. The absence of this type of model has led to the lack of evaluation on Indonesian tax reform and tax policies not achieving their objectives. Since its initial start in 1983, the total tax ratio to GDP and the compliance rate indicates minor improvement. Indonesian tax collection, tax administration and tax compliance still need a long and difficult effort to reach a stage of compliance and collection comparable with more advanced countries.

Another aim of this thesis is to analyse the different tax contributions from different levels of income and how tax policy would affect this distribution of income. To do so, the study analyses the relevant connection between high income contributions and Indonesian income distribution. The study also estimates the potential revenue impact, analyses the winners and losers from the policy, and examines the impact of the reform on income distribution. This is especially because reform and policy changes will create winners and losers. The winners are individuals or firms who experience less of a tax burden after reform, and vice versa. Effectiveness and efficiency, neutrality, vertical and horizontal equity and trade-offs among them must also be recognized and appropriately addressed (Khalilzadeh-Shirazi & Shah, 1991; Burges and Stern, 1993).
1.2.1 General Research Questions

This thesis applies a static microsimulation model to represent the present structure of the Indonesian PIT. This was applied to the base file built from a combination of available Indonesian income survey and administrative micro datasets. Therefore the major research question that can be addressed is “How to build a personal income tax microsimulation model for Indonesia that can assess the full distribution of the Indonesian personal income?”

1.2.2 Specific Research Questions

Eventually, this question will lead to five specific research questions:

1. How to build a personal income tax microsimulation model with Indonesian databases?
2. How can matching and imputation procedures reconstruct Indonesian micro data to capture the full distribution of the Indonesian personal income?
3. What type of microsimulation model can be built to analyse Indonesian Personal Income Tax?
4. What impact will tax reform in Indonesia have on potential tax revenue and on the distribution of the tax burden across disaggregated tax payers and the tax gap?
5. How have the changes in income tax legislation in Indonesia, more specifically the 2008 reform, affected the potential revenue from personal income tax, the distribution of the tax burden, and the efficiency of the Indonesian income tax collection?
1.2.3 Contributions

As the main contribution, this thesis builds a personal income tax microsimulation model for Indonesia. Previous studies of income tax reform, its revenue, and the distribution of income tax burden, have used micro-level tax administrative data and survey data of several developed countries. This has been done in Australia, Canada, Denmark, Germany, Sweden, the US, UK, Italy, Spain (Creedy & Kalb, 2006; Buddelmeyer, Creedy & Kalb, 2007; Baekgaard, Klevmarken & Olovsson, 1996; Eason, 2000; Mc Grath, 2000; McCann, 2000; Naylor, 2000; Kitcher, 2000; Aaberge et al., 1999; Granell-Perez, Fuenmayor-Fernandez & Higon-Tamarit, 2006). However there are still too few studies that have used a microsimulation model for developing and transitional countries. The reason may lie in data requirements and lack of modelling resources. Previous tax microsimulation-based studies in developing, transitional and less industrialized countries include examples from Russia (McNab & Wallace, 2000), the Czech Republic (Stepankova, 2002), and Jamaica (Alm & Wallace, 2007). So far there has been very limited analysis using microsimulation to analyse the distributional and revenue impact of income tax reform in Indonesia. This current empirical analysis fills this gap.

This microsimulation approach is the third attempt to analyse Indonesian tax performance. Previous Indonesian studies include Marks (2003), who analysed the potential PIT using the 2002 Susenas, and Yuwono (2008), who built a PIT microsimulation using administrative tax return data as the base file. This thesis will combine the two approaches by using both Susenas and administrative data in addition to another national survey called Sakernas. Improving the Indonesian personal income tax base file for use in a microsimulation model is
one of the main focuses of this study. The improved base file will contribute a static PIT which will be more reliable to answer the main research questions of the thesis. In addition, the built microsimulation model will also be a powerful input for the Indonesian tax authority for analysing the impact of policy reform.

The contribution of this thesis goes beyond developing this microsimulation model. To start answering the research questions, an accurate and detail PIT base file was constructed. This was required for the estimation of the potential income and the number of potential tax payers in our further analysis. The base file was sourced from several surveys which were available from the Indonesian Bureau of Statistics (BPS); namely the National Socio-Economic Survey (Susenas) and the National Labour Force Survey (Sakernas). Those two sources were statistically matched. Then to augment the limited coverage of high income sample of the survey data, high income earners from personal tax file returns were imputed into the base file.

### 1.3 Review of Literature and Data

#### 1.3.1 Microsimulation Modelling

Microsimulation is the use of micro data about a person, household or firm to analyse the impact of socio economic changes on each individual unit, and to give an overview of the distribution impact on those individual units and to aggregate the impact (Mitton, Sutherland & Weeks, 2000). Since it started in the 1960s, the microsimulation approach has continued
to be developed and enhanced by the rapid development of computational technology. Microsimulation has proven to be a powerful tool to estimate the revenue and distribution impacts of changes in a tax policy - especially in developed countries (Gupta & Kapur, 2000). Its distinct feature is the ability to estimate both aggregate revenue change and detailed changes in the distribution of each unit of individual tax payers. The estimation is built by replicating all the tax codes from the provision of old tax laws and new tax laws in each of the individual units in the dataset.

There are two types of microsimulation models for tax and transfer that have been mostly used by academia and government. Static microsimulation models are used to examine potential short-term impacts of the detailed policy changes to tax and transfer, and dynamic microsimulation models are used to examine long-term behavioural impacts of the changes, particularly on labour participation and on retirement issues (Lewis & Michel, 1990).

Static microsimulation models need cross-sectional information on individuals, households or firms at a certain point in time in the base year to answer the policy question. Microsimulation begins with micro data. The time lag between the frequency of data collection and its publication must be taken into consideration. For example, data on income and expenditure from the Indonesian socio economic survey is collected and published every three years. Depending on the objective of the analysis it might need to be aged to certain periods required for the analysis.
Harding (1996) explains that ageing techniques are critical features that differentiate static and dynamic models. Static microsimulation models use two basic steps for data ageing; reweighting and uprating. Reweighting, sometimes also called ‘grossing up’, involves adjusting the weight of each individual in the household in the dataset. Reweighting usually ages the sample to bring it up to date or up to less than three years in the future. Uprating relates to monetary value adjustments to reflect fluctuations from the original date of the survey or a forecast of future fluctuations. Some datasets might not need ageing as they will be applied for analysis within the same period that the data was collected.

After ageing, a set of computational programs adopting the same rules as the government tax policy are applied to each of the individuals in the dataset. The result of the set of present rules is compared to the set of future scenarios. The end result of the static microsimulation is a measure of the immediate impact of the policy change or the “morning after impact” with the assumption that there is no behavioural response to the tax policy changes (i.e. the individual did not change its consumption or working patterns due to the new tax arrangement) from each of the individuals in the dataset.

The dynamic micro simulation model needs more detailed information for the micro dataset, and cross-section data generally is not sufficient for a dynamic model. The dynamic model needs more comprehensive longitudinal or panel surveys to capture two major issues in individual behaviour estimation; heterogeneity and time dependence (Heckman, 1981). Dynamic microsimulation is more expensive and time-consuming to construct, as it usually includes behavioural responses and forecasts the effects of a change in government policy on the population or a sample of individuals, households, or firms.
Ageing in a dynamic model follows the change of each attribute in each micro unit through time. Harding (1996) further explains that the change in the attributes follows the probabilities of major life events, such as death, marriage, divorce, fertility, education, and labour force participation, for each of the individual in the dataset. To sum up, dynamic microsimulation will recalculate an individual’s attributes for each time in the simulation period.

There have been many other developments in tax-transfer microsimulation. Some studies from Ahmed and O’Donoghue (2007), Davies (2009), Feltenstein, Lopes, Porras-Mendoza, and Wallace (2013) surveyed the history of microsimulation for developed, transition and developing economies. Since Guy Orcutt the founding father of microsimulation started his work in the late 1950’s, Pechman (1965) initially built the individual income tax model in US, Bossons (1967) for Canada. Schulz (1968) then made a model for retirement status of labour force. Orcutt with colleagues (1976) improved the efficiency of microsimulation exercises with the development of DYNASIM. Other developed countries (Australia and some European countries) have also applied and developed extensions to microsimulation. The arithmetical or accounting model was further applied to study the distribution impact of indirect taxation to specific budget for certain commodities (Decoster, 1995).

Behavioural microsimulation models have also improved with the incorporation of labour supply and leisure (Creedy and Duncan, 2002). The dynamic microsimulation model has been developed to follow the structural changes over time by aging the population. This kind of study required an improvement in the availability of panel data for the input. These kind of
dynamic microsimulation models have also been built to study other social sectors such as education (Grimm, 2005), health (Cogneau, Grimm and Robilliard, 2003) and spatial analysis (Ballas, Clarke, Dorling, Eyre and Rossiter, 2005). In the next chapters more detailed information on the improvement in both static and dynamic microsimulation will be elaborated.

Another development in microsimulation modelling is to estimate the impact of the simulation on overall macroeconomic conditions such as economic output or Gross Domestic Product (GDP), growth, inflation or employment (Bourguignon, F., Bussolo, M., & Cockburn, J. (2010). Computable General Equilibrium (CGE) is one tool that can be used to link with a microsimulation model mainly through the impact either models have on a specific industry of employment or occupation (Davies 2009; Buddelmeyer, H., Hérault, N., Kalb, G., and van Zijll de Jong, M. 2012; Vidyattama, Y., Rao, M., Mohanty, I., and Tanton, R. 2014). Ahmed and O’Donoghue (2007), Davies (2009), and Feltenstein et al (2013) have discussed the strengths and weaknesses of the CGE model compared to microsimulation as well as the potential to link the two models to simulate the impact of microeconomic changes on macroeconomic conditions and vice versa. Nevertheless, the macroeconomic impact is outside the scope of this thesis and therefore we will not discuss this link in this thesis.

According to Mitton et al., (2000) the preferred model will depend much on the policy question to be addressed. Based on the current condition of the data available and the surrounding debate about the Indonesian personal income data quality, this present study emphasizes the contribution of a better approach to the construction of the Indonesian income base file. This is where this thesis will contribute to the literature as section one and two of
chapter 2 on the Indonesian income data will discuss more the issues surrounding the
Indonesian personal income data quality and how this thesis will try to solve these issues.

This study utilises all of the available data sources to construct an Indonesian full-income
distribution to improve the microsimulation output. As we will show in chapter 2, by
capturing the full income distribution, the coverage of incomes used will be improved. The
ability to simulate potential tax liabilities to match the real total tax revenue will also be
improved. In addition, this thesis will allow an assessment of the distribution impact on the
full income distribution (as stipulated by the Indonesian income tax law amendment). More
specifically the simulation will be able to cover the impact of tax policy change to the
highest-income decile (wealthy and super-wealthy tax payers). Later in chapter 3, a static
microsimulation model will be built to address the main research questions about the impact
on the potential tax base, revenue and distributional tax burden of the 2008 tax policy
changes.

1.3.2 Application in Different Countries

Microsimulation has gained recognition as a powerful tool to examine the distributional
impact of some underlying public policies. Its capacity to perform simulation on the
condition of individual units before and after the policy implementation has made this tool
popular in several countries. In some developed countries, the models have been intensively
developed for more than three decades. Thus there are many studies on the application of the
microsimulation model to analyse tax-benefit reform. To support the preliminary construction
of the Indonesian PIT microsimulation model, this thesis focuses more on studies that evaluate key development areas and key learning processes of tax microsimulation modelling. The highlighted keys are the main considerations for guiding the preliminary step in building a reliable model.

Atkinson, King & Sutherland (1983), Citro & Hanushek (1991), Eason (1996), Naylor (2000), Wagenhals (2004), Granell-Perez, et al. (2006), Lloyd (2007), among others described, summarized and evaluated recently used tax-benefit microsimulation models in the United Kingdom, the United States, Canada, Germany, Spain and Australia. The range of studies covered the implementation in Europe, North America and Australia. Some of them advocated the use of a survey as the base file for the microsimulation model. The argument was based on the wider range of information and coverage of surveys that would enable them to increase the range of analysis. On the other hand, some emphasize the detail of the analysis using anonymous and weighted tax return data to replicate the tax code more precisely. Some tried to use a hypothetical family, either based on the information from surveys or administrative tax return data, in order to build a hypothetical model with the aim of testing the hypothetical tax-transfer policy scenario. Lastly, some advocated a comprehensive database by statistical matching and imputation of different data sets. These were proposed to overcome the limitation and weaknesses found in both survey and tax administrative data. They argued that tax administrative data suffered from a lack of coverage of non-tax payers and/or low income households whereas survey data had less detailed income data. In the present study, imputation and statistical matching was applied to increase the completeness of the sample; in the case of missing variables or value of the observed variables. The imputation included estimated values or variables to substitute for the missing items without significant distortion to the microsimulation model. The statistical matching can be
accomplished if there are different sources available for the related incomplete variables. Complete data enhances the accuracy and enables wider range of the analysis.

After several methods of data matching, some studies urged for the reconciliation of the final result with a valid benchmark. Usually the benchmark will be other available aggregate data of actual amount of tax collected and other accurate sources of the input information for the estimates of national accounts data.

Forward projection was needed to provide advice about future tax yields and the future cost of making the proposed tax changes. Some researchers are concerned about the necessity to undertake behavioural modelling when tax reform was specifically aimed at influencing tax payer behaviour. The development of microsimulation as a policy analysis tool has involved academia, government (usually on behalf of the Inland Revenue Authority and Ministry of Finance or the Treasury) and the public through the congress (like in US) or parliament (like in Australia, Indonesia and several others). The need to make well documented and more transparent models was also endorsed by others.

An important milestone in microsimulation models was the development of EUROMOD, a European Union wide microsimulation model to enable cross country comparisons. It integrated fifteen EU countries and aimed to overcome the difference datasets and modelling assumptions to enable analysis at the national level and multi-country EU level (O’Donoghue, Sutherland & Utili, 1999).
Other than the evaluation of tax-benefit microsimulation models for those which are already in their mature stage of modelling experience, there are some other studies and applications of microsimulation in other developing and transitional countries. Some of the examples are Vecernik and Stepánkova (2002), Alm and Wallace (2007), Bargain, Morawski, Myck and Socha (2007), Wilkinson (2009) among others conducted empirical tax-transfer microsimulation studies for the Czech Republic, Jamaica, Poland and South Africa. In addition, Lelkes (2007) conducted a survey about the experience in tax-transfer microsimulation of ten countries which joined European Union in 2004.

From Lelkes (2007) survey, it is noted that four out of the ten countries, the Czech Republic, Hungary, Estonia and Slovenia, have built national models. In addition to that, Bargain et al. (2007) succeeded with their first fully functional Polish model. The majority of the existing models are static with the exclusion of the existing Estonian model, which included behavioural responses. Most of the models were still under-developed and recently built with the exception of the Hungarian model (TARKI). It has been available since 1995 but since then it has not been used much for policy analysis.

A Jamaican model was built by Alm & Wallace (2007) to address vertical and horizontal equity in Jamaica’s labour income tax. Another African model, the SAMOD-South African microsimulation model was built in 2009. It used the EUROMOD platform and aimed to encourage and to expand the use of microsimulation as an evidenced-based policy analysis tool. It augmented the previous microsimulation model in both academia and government communities in South Africa.
Most models use a survey of income and expenditure as their base file; only TARKI, the Hungarian model, was built under matched income surveys of expenditure and personal tax records. Some possible improvement through data matching and imputation were of concern and some of the researchers advised that the results of their microsimulation needed further verification using better data sources. Advancement of microsimulation model development has been relatively slow due to low political interest in distributional policy analysis, low research funding and research capacity within the internal government users especially, and poor quality of, and limited access to, micro-data for some countries. However the urgency to produce fair and equitable tax policy as well as government's need to maximise tax revenues have spurred the issue of the availability of reliable data.

As previously mentioned, this is not the first time a microsimulation technique has been used in studying the Indonesian tax policy. There are at least two examples of tax microsimulation in Indonesia. One of those studies is a doctoral dissertation by Yuwono (2008). Yuwono used a sample of micro-level data from salary-income withholding tax returns from the Indonesian Directorate General of Taxes (DGT). She estimated the taxable income elasticity using differences in difference econometrics and built a theoretical model of labour response behaviour following income tax change using a basic labour-leisure model. In the last part, using microsimulation with administrative data, she empirically examined the distribution of the income tax burden across different income groups and the government’s withholding income tax collection from some proposed scenarios of personal income tax law. Her microsimulation results suggest that there is a trade-off between revenue impact and the distribution of the tax burden. If the choice of income tax reform had a broader impact on the
distribution of tax burden to all income groups then revenue loss will be compensated. On the other hand if the larger burden was targeted to a high income group then revenue yield could be expected.

Another unpublished study was conducted by Marks (2003) as a part of the USAID project with the Indonesian Bureau of National Planning (BAPPENAS). This study calculated all the income and income tax obligations in Indonesian households using Susenas 2002. It conducted the simulation based on the applicable Law No. 17/2000. In this simulation, based on Susenas’ survey unit, the household was used as the unit of analysis whereby as stipulated by the law, the Indonesian income tax unit is an individual. The individual tax payer, based on status of marriage, could claim non-taxable income from its marriage and its eligible dependent of maximum 3 (three) in one family. Marks’ approach highlights a big gap between potential revenue and real collected revenue from income tax (article 21) for salary earners, from personal business (article 25), final tax for rent and/or some financial and capital income, and from interest (article 23), dividends, royalties and prizes. Mark’s approach was one of the initial detailed approaches used to assess Indonesian personal income tax using survey data.

This thesis is built on these two previous Indonesian studies with some key factors learnt from other countries’ studies that were taken into consideration in this present study. The research questions that were raised from these two Indonesian studies are answered by developing an improved base file and more detailed microsimulation for Indonesian personal taxpayers. We will employ all available income micro data to create the best possible base
file for assessing the distribution of Indonesian personal income, and this base file will then be used to build a personal income tax microsimulation model for Indonesia.

1.4 The Available Data Source

A microsimulation model begins with micro data, which can be in the form of administrative data or survey data. Due to the nature of data collection, both survey data and administrative data have their own inadequacies. Inadequacies in the survey data can be due to some missing variables, under reporting or misreporting of income and expenditure, and non-response to the survey. Our own preliminary observation indicates that Indonesian survey data suffers a lot from under-reporting of income and under-reporting of high income respondents. Based on the discussion with an officer from the Indonesian bureau of statistics (BPS), this is caused by the voluntary nature of respondent participation and by the limited access of enumerators to several luxurious housing complexes (BPS, 2010). Indonesian administrative data particularly suffers from being non-weighted data. It also suffers from incomplete coverage of taxpayers due to low submission of taxpayers’ returns. We also found some missing values and variables due to manual recording errors.

In anticipation of the data problems, modellers are often required to adjust by using various techniques like data matching (Paas, 1986). There are three major data sources for building a microsimulation of Indonesian Income Taxes. This study attempted to solve data problems by combining each of the available micro data using statistical matching and imputation. The main base data for this analysis was derived from the Indonesian Social Economic Household
Survey, Susenas 2008. This Susenas was combined with the Indonesian Labour Force Survey, Sakernas 2008 especially for labour-income completeness and for the weighting process. Lastly, DGT administrative data from 2008 of reported personal income tax returns augmented and eliminated limitation in the survey data.

1.5 Model Construction Strategy

The first step was preparation of the base file. The chosen base file was the Susenas 2008 micro data. It was chosen since it has a wider representation than Sakernas 2008, especially its coverage of the Indonesian population and Indonesian economy. So far, there is no comparable survey that matches the coverage of Susenas both in its variety of income data and household characteristics information. Susenas is also a base file for calculating Indonesian aggregate economic indicators (BPS, 2010). On the other hand, Susenas is not designed for simulating tax policy. It can be expected that the income component of this survey is still less detailed compared to the tax administrative data. Hence, as argued by Wagenhals (2004), this type of income information might be suitable only for rough simulation.

Susenas 2008 includes all household income information except information on labour income. Based on the employment variable, the missing Susenas labour income data was imputed by matching the labour income variable from Sakernas. The Susenas-Sakernas base file was then developed as an imputation base file using a modified cold-deck method. A combination of the 12 most important variables were used to find similar individuals in the
recipient database. This similar individual then had their wages imputed from the donor database. Those chosen variables were province, municipality, district, sub district, main occupation sector, highest education level, sex, age, marital status, number of household members and dependents. The complete methodology and process will be discussed in chapter II.

The Susenas-Sakernas base file was then enhanced with the rich administrative data. This was done by imputing income from the administrative data to augment the Susenas-Sakernas base file. We firstly grouped the cells based on 17 predetermined income brackets, and then imputed all of the available income information from the administrative data to the Susenas-Sakernas base file. The imputation was done based on several geographic and socio-economic matching criteria. However we didn’t reweight the administrative data. We added new observations into the Susenas-Sakernas base file. All observations imputed from the administrative data were weighted as one individual. This enabled evaluation of the possibility of imputing observations of high-income groups and of other detailed variables from the administrative data to the survey base file. This process provided a better potential income base file as an input for the microsimulation model. When the potential income base file was finished, non-taxable income was estimated. From this, a gross-up equation was built to identify individuals who earn gross salaries above a taxable income threshold under both previous Law no 17(2000) and the latest Law no 36(2008). This became the final base file of an original parameter dataset according to Law no 17(2000) and a changed parameter dataset according to Law no 36(2008).
The second component of the model is a parameter dataset which details the Indonesian tax structure variables and mimics the Indonesian tax code in tax file returns. It contains information on tax rates, exemptions, tax credits and other main components for income tax calculation as stipulated by the tax law. To sum up, the original parameter is the parameter of the current law which for this thesis contained the dataset based on the previous law. Parameters were changed to reflect changes from the income tax reform, and particularly for this thesis, the dataset based on the new law. These parameter datasets were used as input data for the tax calculator. The calculator calculates potential revenue, the tax gap between potential and real revenue, examines distribution of tax burden, and provides a winner-losers analysis and compliance analysis. The results for both parameters were disaggregated by income groups, occupation/business sectors and geographical-tax office’s services area. The complete methodology and process will be discussed in chapter III.

1.6 Thesis Outline

Based on its intended contribution, the thesis is organised as follows:

I. Introduction

- This chapter presents the background of the study, its scope, the research questions and the hypothesis leading to the structure of the thesis.

- Observation of the evolution of tax microsimulation modeling as tax policy analysis tools and its implementation across countries.

- General summaries of how the research was conducted.
II. **Income in Indonesian Micro Data**

- This chapter studies the Indonesian micro data (both survey and administrative data); its strength, weaknesses, and the literature on empirical studies using each micro data.
- The chapter presents the detailed process of the base file construction. This includes data matching, imputation and the final merging process, up rating techniques, forward projections and data validation.
- The role of better micro data to capture full income distribution and support microsimulation model for Indonesian personal income tax.

III. **Developing Indonesian Labour Income Tax Microsimulation**

- Overview of the development of the tax microsimulation model
- Overview of the Indonesian experience in tax modelling and the possible microsimulation contribution to the Indonesian model.
- The construction of the income tax microsimulation model
- The possible applications for current and future policy simulation

IV. **The Revenue and Distribution Impact Analysis of Indonesian Personal Income Tax Reform in 2008**

- The development of a static microsimulation model for Indonesian potential personal income tax.
- The Impact of the 2008 Indonesian Personal Income Tax Reform on People’s Income.
• The impact of the latest 2008 individual income tax reform on the estimated potential tax gap across categories of tax payers, and a detailed discussion and analysis of the Indonesian government tax revenue enhancement program.

V. Conclusion

• This chapter relates the empirical results and each final discussion from chapters II to IV.
CHAPTER II. INCOME IN INDONESIAN MICRO DATA

2.1 Introduction

For some developing countries, income and expenditure data collection by using a household survey is not a new phenomenon (Deaton, 1997). Indonesia is among the developing countries which have relatively better income micro data sourced from continuous and long established surveys carried out by the Bureau of Statistics (BPS). Since the 1960’s, BPS household surveys have provided a rich source to study a connection between individual economic behaviour and government policy intervention.

Currently, the Indonesian economy is among the three fastest growing economies of the G-20 countries. It constantly records higher and steadier growth compared to other G-20 members. It still grew when the 2008 global financial crisis hit the world’s economy. Since 2004 Indonesia has gradually shifted to become a middle-income country with an annual Gross National Income (GNI 2004) per capita of USD 1,070.

Several studies and official government statistics noted a long history of constant and moderate Indonesian inequality. Since 1964 and up until the recent data for this study (2008), the trend of the per capita household consumption’s gini coefficient ranged from as low as 0.32 and was never greater than 0.40 (BPS, 2009; Mishra, 2009; World Bank, 2006; Yusuf, 2006). Yet the proportion of the Indonesian population living on only $2 per day is as high as
the lowest income country in the East Asia region. Indonesia has 49% of its population with an income under US $2 per day while the other low-income countries, like Cambodia, Laos, Papua New Guinea and Vietnam, recorded an average of 53.8% (World Bank 2006:24).

One of the major issues in the estimation of the lower and higher incomes in Indonesia is the unavailability of good income data. This issue has been an ongoing problem in the study of the Indonesian income distribution (Cameron, 2002; Leigh & van der Eng, 2009). As mentioned above, the most notable concern is its impact on the study of people with low income and the level of poverty. Most of the studies on poverty used consumption expenditure data to proxy the income in Indonesia. The availability of good income micro-data is not only a crucial starting point to better understand the structure of Indonesian income but also to understand the impact of government policy intervention, such as their tax-benefit policy reform, on income redistribution.

Hence, there is a challenge to improve the income database while using the available data sources. It is especially challenging to prepare a better database for our income tax microsimulation model.

This chapter aims to explore and assess Indonesian income data and apply a better method for building the micro data with more reliable income variables for Indonesian microsimulation modelling.
2.1.1 The need of Income for modelling tax microsimulation

Since the first reforms started in 1983, tax laws and regulations, and tax administrations and systems have continuously been improved by the Indonesian government. There are two government institutions who bear responsibility for the improvement. The Fiscal Policy Office (FPO) is the authoritative body for tax policy formulation while the Directorate General of Taxes (DGT) is responsible for tax administration. The main aim of the reform is to source a sustainable level of public financing. The Government has had to cope with increasing tax revenue and compliance target as mandated by the parliament in the national state budget.

These sets of tax reforms had a big impact on Indonesia. Yet there were rarely tools for detailed analysis to assess policy consequences of the increasing target of tax revenue within the current state of tax compliance on Indonesian income distribution or details on the distribution of the impact of current and future tax structures and tax payers’ demographics. The assessment may also have an impact on how the current Indonesian income distribution and the demographics of tax payer could actually drive the Government’s strategy to increase tax revenue and compliance. Learning from tax modelling experience in several advanced economies like the US, Europe and Australia, Indonesia needs a tax microsimulation model. This model could deal with the analysis of the detailed impacts of reforms on the individual decision making unit (Gupta and Kapur (Eds.), 2000; Harding (Eds.), 1996; Mitton, Sutherland & Weeks (Eds.), 2000; Orcutt, Merz & Quinke (Eds.), 1986). Currently the official main model used for proposing changes of tax policy to the parliament is a macro
aggregate model (Ministry of Finance, 2009). In view of the need to prepare future analysis to capture the interaction between policy changes and the socio-economic impact on the Indonesian taxpayers, a proper and robust microsimulation model is needed. The long standing question about the Indonesian tax payer compliance could also be analysed by the model. So far there is no information available on any studies by the DGT administration regarding the level of non-compliance or the outcome of audits. Hence this thesis is not only about providing an answer to the impact of Indonesian tax policy and administration change, but also building and preparing the tools for that.

In the case of tax reforms incidence analysis, microsimulation is often the preferred methodology compared to other methodologies that focus on macroeconomic conditions such as CGE models (Feltenstein, Lopes, Porras-Mendoza, and Wallace 2013). The distinct feature of microsimulation models is that they in provide detailed behavior of individual households and firms. Using microsimulation, the disaggregated details of the individual agents are observed in two types of direct applications (Bourguignon and Spadaro 2006). First, under a highly disaggregated sample, it is simpler to identify the likely winners and losers of a reform. Second, using the disaggregated individual analysis, the results can easily be aggregated to provide a more accurate evaluation of the aggregate financial cost or benefits from a tax reform.

The availability of good income micro-data is an essential ingredient of a good microsimulation model. The micro data on persons or households or other micro units is needed for the simulation of the impact of a policy change on each of the individual units. Microsimulation has a distinct feature compared to other economic, statistical or descriptive
models. It exploits the individual information from the micro base file in each and every step of the analysis. Thus reliability of the micro data that contains the individual information will determine the quality of the built models (Lau, Yotopoulaos, Chou & Lin, 1981; Mitton, et al. 2000; Spadaro 2007). Davies (2009) emphasized that no single-survey nor administrative data comes with the perfect micro data required. The best data available dataset will be used as a starting point (called the host dataset). This host dataset will then be imputed and augmented with values of variables from other surveys (income and or expenditure surveys) and administrative data (tax returns, social security). This augmentation and imputation should be reconciled with the national macro data from independent sources (e.g. national accounts).

2.2 Indonesian Income Micro data

In Indonesia, National Social and Economic Survey (Susenas) data would be one of the main datasets to source Indonesian income. Susenas is an official survey by the Indonesian Bureau of Statistics (BPS). It has a long history and is one of the oldest regular national surveys used to collect the Indonesian social and economic indicators. The other source of income information from employment is provided by the National Labour Force Survey (Sakernas). It sources national labour market characteristics from the working age individuals in the sampled households. It includes the information of labour-income (in kind and in cash salary/wages). The survey was initiated by BPS in 1976, thirteen years after the commencement of Susenas (RAND, 2010).
The last source of income data is from personal income tax returns administered by the Directorate General of Taxes (DGT). DGT is the Indonesian tax authority under the Ministry of Finance (MoF). It includes pooled cross-section micro-level data that contains detailed information on taxpayers’ income. DGT administrative data of reported personal income tax returns will augment the survey data. However, the coverage of this data is limited to those who have submitted their tax return. In addition, this data has also had limited usage due to the difficulty in accessing it. As stipulated in article 34 of the Indonesian income tax law 2008, these tax payer data are confidential. Therefore, there are not many research projects that can maximise the use of this data.

This section will look closely at these three main datasets and the difference between them to give an indication of how these different datasets can be used to fill the gap in the Indonesian micro-data.

2.2.1 SUSENAS

History of the survey

Susenas is a national survey to collect social and economic indicators. This survey is conducted through face to face interviews between enumerators and selected respondents. Susenas has two sets of questionnaires, core and module, for individuals and households respectively. A core questionnaire was first introduced in 1963 and has been conducted annually since 1989. The Susenas-module was then initiated in 1992 to cover more detail and
to provide a specific questionnaire on income and consumption, welfare, health, social, culture and education.

The Susenas-core is focused on capturing necessary information which may change in the short (yearly) term. The Susenas-module captures specific information which may not change much in the longer term (more permanent). The Susenas-module collects specific information every year over three year periods. Within the three yearly periods, specific information on household consumption expenditure and income is collected in the first year. The second year is for the household welfare, social culture, travel and criminality module; then the health, nutrition, education and housing module is in the third year. So the Susenas-module comprises of these three modules whereby each is added in a three-year cycle. Started in 2003, a subset of core-module samples which comprised of several households was selected for their consumption and income information to be included as a panel sample.

**Survey methodology**

Susenas uses a stratified multi stage cluster sampling method for the data collection. Two strata (urban and rural areas) are used for each municipality. For each municipality, Susenas applied a 2-stage cluster sampling method to urban areas, and a 3-stage cluster sampling design for rural areas. Urban areas are divided into census blocks (CBs) which comprise approximately 100 households. The first stage of the sampling involves choosing a number of CBs using linear systematic sampling. Then, from each selected CBs, by using linear systematic sampling, sixteen households are selected for interview. While for rural areas, by using the ‘proportional to size’ principle, a number of districts (Kecamatan) are selected. It is
meant to maintain the probability of each sub-district being selected to be proportional to its size (i.e. number of households). In the second stage, from each selected sub-district a number of CBs are selected. Lastly, by using linear systematic sampling, sixteen households are selected from each of these selected CBs.

**Questionnaire and variables**

Information in the questionnaire is collected through direct, face to face interview between the enumerator and the respondent. All individual questions are supposed to be directly asked of all people 15 and over in the household. Questions about household information are asked of the head of household. If the head of household is unavailable then the wife/husband of the head of the household is asked, and if they are not available then another household member who knows the characteristics of the information is asked. The household survey questionnaire covers:

- size of household;
- type of housing and amenities;
- household income and expenditure;
- employment status,
- form of transport used by the household members;
- telephone and computer usage and;
- question about business loans.

The survey for persons covers:

- sex;
- age;
- marital status;
- crime;
- school participation;
- health;
- immunization;
- literacy and numeracy;
- telephone and computer usage;
- daily activities;
- employment status and;
- contraceptive use.

Population weightings are included for both surveys. This weighting is based on the ‘proportional to size’ principle that was applied in the survey design.

**Coverage**

When it first started in 1963, the coverage was limited to Java Island with just 16,000 households. Gradually (1964, 1967, 1969, 1976, 1978, 1979, 1981, 1989 and 1992) Susenas were conducted with larger samples and a larger coverage area until it finally covered all Indonesian areas. In 1993, when Susenas commenced core and module data collection, the household sample varied between 202,500 in 1993 to 278,352 households in 2006. The sample increased to 285,904 households in both the 2007 and 2008 surveys.
2.2.2 SAKERNAS

History of the survey

Sakernas was initiated in 1976 to cover the national labour market and work force data and, especially, updated characteristics of all working age individuals in the sampled households. It is mainly designed to monitor the general dynamics of the labour force and the change in its structure between the enumeration periods. It provides periodical snapshots of the Indonesian workforce by updating data on the number of employed and unemployed, and individuals who used to be in the work force, are seeking employment or moving from one employer to another at the regional and national level.

The survey has generally been conducted on an annual basis since 1986, drawing on either quarterly, annual or semi-annual observations. Sakernas 1986-2004, was conducted annually with some quarterly trials from 2002-2004. The trials covered all provinces except Papua, Maluku and Aceh. From 2005-2010, Sakernas was conducted twice a year (in February and August). Since 2011, the survey has been conducted quarterly as input for the Early Warning System (EWS) for the Indonesian employment sector.

Survey methodology

Sakernas enumerators visit the sample households and collect information by using direct interview and face to face methods. The survey applies a two-stage sampling design. Census Blocks are the first stage sampling unit (or primary sampling unit), and households are the
second stage or ultimate sampling unit. The sample frame is a master file of villages which are divided in census blocks with probability proportional to size. Each census block consists of around 80-120 households. From 2007-2010, samples were updated by using a rotation pattern with 25 percent of samples changed every semester. From 2011-2015 a different rotation pattern applied with 25 percent of samples changed quarterly instead of bi-annually.

**Questionnaire and variables**

Sakernas has evolved with some changes both in its household and in its regional sample coverage. From the selected households, general information from each household member (name, relationships to head of household, sex and age) were collected. Then the enumerator collects information about marriage status, education and employment for members of the household who are 10 years of age and over.

In detail, there are two kinds of information collected in the Sakernas questionnaires; namely household and individual information which is asked for persons aged 10 years and over. Data on individual information includes:

- member of household information;
- name;
- relation to the head of household;
- sex;
- age;
- marital status and;
• educational attainment.

The individual information includes:

• activities during last week:
  o worked;
  o temporary not working;
  o looking for work;
  o attending school and;
  o doing housekeeping, and others (such as pension, disabled).

• Further, for those who are working:
  o place of work/industry;
  o employment status;
  o total hours of work during last week and;
  o total wage/salary received.

• Finally, for those who were looking for work:
  o the duration of looking for work;
  o were they looking for full or part-time jobs and;
  o how they were looking for work.

BPS adopted International Labour Organisation (ILO) standard for Sakernas. BPS has made several changes to the unemployment concept. For Sakernas 1986-2000, the unemployed were defined as those who do not work and were looking for a job, while for Sakernas 2001-present, unemployed are those who do not work but a) currently looking for a job, or b) preparing their business, or c) feel it is impossible to find a job, and d) have a job but have
not yet started. There were also changes in the occupational status. Five categories became 7 categories, namely: 1) Self-employed; 2) Business with non-permanent employee/unpaid employee; 3) Business with permanent/paid employee; 4) Labour/worker/employee; 5) Independent worker in Agriculture; 6) Independent worker in Non-Agriculture; 7) Unpaid employee.

**Coverage**

When it first started in 1976, the Sakernas sample size was 95,400 households from all provinces except East Timor. Gradually, from 1978 to the present, Sakernas started sampling annually with different samples and coverage. The Sakernas set used for this present analysis was gathered in August 2008 and covered 293,088 households and 931,890 individuals from all provinces in Indonesia. The large number of samples of Sakernas 2008 makes it possible for the data to be analysed at the district level.

Besides increasing the geographical coverage, the coverage was also increased to include all industries and occupations. A lower age limit of 10 years was applied in the field enumeration as provided in questionnaires, although an applied lower limit of the working age is 15 years of age. Topical coverage on data was provided by industry, occupation, status in employment, age, sex, hours of work, level of education, and region.
2.2.3 Administrative Data/Income Tax Return

This administrative micro data from the Indonesian tax administration systems can only be accessed by special permit from the DGT. This might be the only administrative income data which is backed up by the Indonesian tax constitution. The Indonesian self-assessment system requires each taxpayer to be responsible for calculating, analysing and reporting their tax liabilities using forms provided by the DGT. The main dataset for this static microsimulation model was derived from three forms:

1) The Personal Income Tax article 25/29 return, SPT 1770, covers personal income from salary/wages, personal business, self-employment, employees with one or more employers, final taxed-income and other income.

2) The Personal Income Tax article 25/29 simple return, SPT 1770s, (personal income from salary/wages from one single source of income). This is a simplified annual tax form for individuals with only one source of income.

3) The Personal Income Tax article 21 return, SPT 1721, (withholding income from salary/wages). This is a withholding tax return form for the employers. It contains details of tax withheld from the employees.

All of these datasets include pooled cross-section micro-level data that contain detailed information on taxpayers’ incomes. One of the issues with this dataset is that there is possible replication of taxpayers across types of returns. In this case, the same individual may have the equivalent of SPT 1721 and SPT 1770 or SPT 1770s. All of this duplication were removed through the data request to the DGT so that we could proceed using SPT 1721, SPT 1770 and SPT 1770s for the analysis.
In order to comply with confidentiality and secrecy of taxpayers’ information as stipulated in Indonesian Income Tax Law article 34, all identifying information was removed to maintain taxpayers’ confidentiality.

Variables

Variables from the tax return mirror the stipulated tax legislation. The variables evolved with changes in the Indonesian tax legislation. The coverage depends on the tax payer compliance. Indonesia uses administrative measures of compliance, which is a percentage of tax payers who submit their tax return from all registered tax payers in Indonesian tax systems.

In detail, there are two groups of variables from the personal income tax return; individual information and income detail for tax calculation. These are described below.

Individual information:

- name;
- tax identification number;
- address;
- marital status;
- sex;
- (foreign) employment status;
- number of dependent and;
• job’s position.

Income detail for income tax calculation:

• Gross income:
  o salary/pension;
  o income tax allowance;
  o overtime, other allowance;
  o ‘honorarium’ or ad hoc/occasional payments and other payments;
  o insurance premiums paid by employer;
  o income in the forms of goods/services;
  o bonuses, gratification, production services and special religious festival allowance.

• Deductions:
  o occupational expenses;
  o pension expenses;
  o pension instalments as a fraction of pension allowance.

• Income tax calculations:
  o net income;
  o previous net income;
  o previous net income for income tax calculation (annually/annualised);
  o non-taxable income;
  o taxable income (annually/annualised);
  o income tax from taxable income (annually/annualised);
  o income tax withheld from previous year;
  o income tax liabilities;
  o income tax borne by government;
- income tax to be withheld;
- income tax withheld/paid previously;
- over-payment/under-payment of income tax.

**Coverage**

Information from three sources of personal income tax return (SPT 1770, 1770s and 1721) were merged for personal income tax modelling. The merged dataset used for this study was the taxation data for the fiscal year 2008. The final administrative record of anonymous personal income taxation return for the fiscal year 2008 contained 4,061,136 individual recorded returns. It includes cross-sectional micro-level data covering all national records from 956 tax offices for the period. It contains detailed information on taxpayers’ salary/wage income (non-wage income is not included). Most importantly, there is no top coding for the top income group.

**2.2.4 Data Comparison (Strengths and Weaknesses from each Micro Data)**

**SUENAS**

The strength of Susenas is in its richness of information. The content of demographic characteristics and socio-economic variables of the sampled households, and of individual members has made Susenas the main data source for most income and/or consumption-based studies. The Indonesian Bureau of Statistics or Badan Pusat Statistik (BPS) officially uses
Susenas to monitor the poverty status of the Indonesian household. Susenas consumption data is used by BPS for their official estimates of Indonesian poverty by applying a basic needs approach. From 1984, the calculation of the number and percentage of the poor population was conducted three yearly using the Susenas consumption module. Commencing in 2003, BPS published the number and percentage of the poor population annually after they started to conduct an annual panel consumption module of Susenas. The BPS approach defines the poor population as the population with an average monthly per-capita expenditure below the poverty line.

The BPS poverty line is technically estimated from a reference population in the Susenas. The reference is 20% of the population just above the temporary poverty line which is estimated from the last period poverty line inflated by the consumer price index. The poverty line is built with two components: food and non-food poverty lines. The food poverty line estimate is the spending required for 2100 calories of food consumption, while the non-food is based on minimum needs for housing, clothes, education and health (BPS 2012:72).

Studies of Indonesian inequality, poverty and income distribution have used Susenas consumption and income modules as major sources of their micro data. So far, to the best of our knowledge, most studies on poverty and inequality in Indonesia preferred consumption compared to income data. It confirmed the weaknesses of the Susenas’ income variable. It is difficult to find studies that used the income variables from Susenas surveys. Most studies question the quality of Susenas income data, and its inferiority compared to expenditure data. Thus they use Susenas-expenditure as a proxy for income. The per-capita or household consumption Gini-based analysis is more frequently quoted and used in either academic,
public or government analyses (Timmer, 2005; World Bank, 2006; Yusuf, 2007; Miranti, 2010; Suryadarma et al., 2005 and 2010).

Nonetheless, among the studies, some used Susenas income micro data for their analysis (Alatas & Bourguignon, 2005; Cameron, 2002; Leigh & van der Eng, 2009). One particular study by Frankema and Marks (2010) used another three indicators, namely wage ratio of unskilled agricultural to GDP per worker, inter-industry manufacturing wage inequality, and urban self-employment, as an alternative approximation for income distribution studies. Lately, Leigh & van der Eng (2009) used all possible sources of Indonesian income distribution to conduct trend analysis for an international comparison of top incomes share. The sources of data were unique and under-utilised income taxation data (1920-1939, 1990-2003) and earned income data from 1982-2004. Their results suggested a long-term stable trend of Indonesian income inequality. They found an increase in the top 10% income share in the early 1920s and early 1930s, while there was a decrease share in the top 1% in the late-1930s. Further, they found a comparable match of an increase in the top 10% income share for the period 1982-2004 compared to the 1930s. However, a more distinct increase was observed in the top 1% share during 1982-2004. They finally showed that their comparison of top income share in Indonesia, Argentina, India, Japan and the United States throughout the twentieth century concluded that the top income share in Indonesia has been higher than in India, broadly comparable to Japan, and somewhat lower than for the United States.

Nugraha and Lewis (2013) explored Susenas household income and income inequality in Indonesia. They assessed the role of non-market income to accurately measure Indonesian household actual income. They found that imputing non-market income had resulted in a
significant impact on the income distribution; the gini coefficient improved from 0.41 to 0.21 (not to mention the significant increase of more than five times, the income share of the poorest decile). They concluded the use of only market income would bias the measures of Indonesian income distribution.

In the field of microsimulation, so far there is only one Susenas-based study (Marks, 2003). Marks built a tax simulation model based on Law no 17(2000) and provided the first detailed approach using survey data. Marks studied the revenue potential and allocation of household burdens of the personal income tax (PIT) system in Indonesia. His analytical framework was a preliminary step to examine the impact on revenue of the change in the PIT administration, system and regulation improvement. He also provided important insights into the analysis of household burdens throughout the income distribution.

The richness of household and individual characteristics in the core, combined with the power of module data on consumption, expenditure, income, health, and education had made Susenas a rich data source. Yet Susenas still needed improvement to maximise the power of analysing income/non-income poverty, income distribution, taxation and other income based-studies. Some scholars (Cameron, 2002; Leigh & Van der Eng, 2009; Mishra, 2009; Yusuf, 2011) note a limitation on the coverage of the poorest households as well as Susenas’ ongoing inability to cover the richest decile of Indonesian households. Susenas suffers a lot from under coverage of high income respondents, under reporting of income, and under enumeration of income data. The main cause of this is the voluntary nature of respondent participation and often difficult access to Indonesia’s luxurious housing complexes (BPS, 2010). In addition, Groves and Couper (1998) observed that high-income households might
be less willing to be involved in such surveys because of their privacy, time and opportunity cost. These limitations of survey data found in the Indonesian survey are generally in line with findings from studies of other advanced countries. Several studies of micro data quality, and assessment from the existing microsimulation modelling experiences for advanced countries like the US (Radner, 1981; Citro & Hanushek, 1991), Germany (Wagenhals, 2004), UK (Eason, 1996), and Canada (Naylor, 2000) also note their concern about limitations of the income survey data.

**SAKERNAS**

As with the Susenas, the strength of Sakernas lies in the detailed demographic and rich socio-economic variables, national-wide survey coverage, long survey history, and improved methodology due to the years that the BPS has been running. Officially, the BPS uses Sakernas to monitor three main indicators: (1) Employment by education, working hours, industrial classification and employment status; (2) Unemployment by different characteristics and efforts to look for work; and (3) Working age population not in the labour force (e.g. in schools, doing housekeeping and others). One of the most important variables from Sakernas for income-based studies is the information on individual monthly salary/wages. This includes salary/wages in cash and in kind which is earned from the main job over the last week. However, while Susenas includes more detailed income information, Sakernas only collects labour income information. On this point, there are few income distribution studies based on Sakernas income micro data. Most studies found are of labour market analysis. Scholars use Sakernas labour income data to highlight issues such as government salaries compared to private sector pay, minimum regional wages, severance
payment, and the welfare of workers. Filmer and Lindauer (2001) studied the estimates of government versus private employee earnings using data from the 1998 Sakernas and 1999 Susenas. They aimed to contest the previous empirical evidence of policy analysts like Smith (1975), Gray (1979) Wirutomo (1991) and the World Bank (2001) that Indonesia’s civil servants are relatively poorly paid compared to their private counterparts. This fact was often used to explain corruption at various levels of government. Filmer and Lindauer’s study showed that the Indonesian situation is not unique. It experiences the same condition as other countries in their problem of government salary compression.

Other studies looked at the impact of labour income on labour market dynamics in Indonesia. Suryahadi, Sumarto & Maxwell (2001) at the SMERU research institute used Sakernas panel data from 1988 to 2000 to study the impact of the change in the minimum wage on earnings and employment. Further, their SMERU research team employed Sakernas data and a small qualitative survey to analyse the minimum wage and its employment effect. They found that a minimum wage benefits some workers who can keep their job and disadvantages others who lose their job. The clear winners are the white-collar workers and the potential losers are those workers most vulnerable to labour market changes (such as females, youth and less educated workers). They suggest it will be more preferable to increase minimum wages in an environment of high economic growth.

Previous studies show comparable income data and labour characteristics quality and consistency of the Sakernas data. Thus we believe the Sakernas labour income variables will be the best source to complete our attempt to improve the Indonesian income base data. Citro & Hanushek (1991) emphasized the role of overlapping survey data to provide a benchmark
for data validation. Therefore, some of the overlapping variables like demography, socio
economic status and, particularly, labour income in Sakernas will also be the best source for
validation of our variables in the microsimulation base file. In addition to the previous use of
Sakernas in labour issues studies, we will exploit the Sakernas labour income data to study
the personal income tax and income distribution.

Administrative Data/Income Tax Return

This administrative data has the advantage of detailed income information from all taxpayers
who submit their tax return and readily available information for the tax calculation based on
the Indonesian income tax law. Details of personal taxpayer information in these datasets are
annual gross income, net income, annual net income, personal exemptions, taxable income,
statutory marginal tax rate, income tax liability, and individual characteristics. The nature of
the administrative data, which is richer in income information and more ready to use, is a key
addition to the power of survey data (Citro & Hanushek, 1991).

However, there are several limitations to using Indonesian tax administrative data including:

- the administrative tax return database has less coverage of the low income earners and
does not cover non-taxpayers;
- the data are not weighted to population totals; and
- a complete pooled cross-section dataset for all registered taxpayers for the observed
  period is not available.

- This last point is due to a low compliance rate on tax return submission. Indonesian’s tax
  payers’ compliance averaged between 35% and 40% before 2008 (DGT, 2010).
There are also some missing values and variables in our tax return data due to manual recording errors. These make it difficult to conduct an accurate analysis of the Indonesian potential tax revenue simulation. We therefore chose to maximize the usefulness of this administrative data by using it to provide information on the top income individuals.

2.3 Filling the Gap in the Income Survey Data

From now, it should be clear that we need to maximise the advantage of each available dataset while, at the same time, minimising the limitations. The aim is to construct a better income base file to enable us to capture the full distribution of income in Indonesia for microsimulation analysis.

Atkinson (1997), in his book on the analysis of household surveys, explained that generally the two sources of income micro data, a household survey and an income tax return, have their own strengths and weaknesses. He suggested the use of both sources for better estimates of the income distribution studies. As a result of the assessment of the advantages and disadvantages of the available survey and income administrative data sources, modellers are often required to use various techniques of data matching. Modifications to the survey and administrative data will be needed so as to adjust them to be suitable for the microsimulation base file (Paas, 1986).
We attempted to maximise the advantages of Susenas, Sakernas and PIT Return data by combining each of the datasets. We used data matching, variable imputation and re-weighting for this exercise. The choice to improve the model base file by combining available survey data and imputing administrative data followed a recommendation from Citro & Hanushek (1991) to combine the use of March CPS (Current Population Survey) and SIPP (Survey of Income and Program Participation) with taxation administrative records data in the analysis of panel evaluation works to improve the US microsimulation model. Other previous US studies by Radner (1981) also emphasized that, even since 1964, the Bureau of Economic Analysis (BEA) of the US Department of Commerce started to combine several income data sources in order to provide better estimates of the income distribution. The work has then been continued by other US income data authorities like the Department of Health and Human Services, the Bureau of Census, and the Office of Research and Statistics at the Social Security Administration. To overcome non-response and missing and/or incomplete variables, they intensively use data matching. Whenever information from the same person is available from another database, an exact match can be made. When exact matching is not possible, they use statistical matching to improve databases for income distribution studies. Statistical matching will look for a similar person (on the bases of similar characteristics) from a different database to be the “donor” for the missing information/variables.

The related approaches were also confirmed with experiences of other advanced countries that have used both the income data from administrative and survey data to study their income distribution. Decoster and Van Camp (1998) attempted to build a relationship between a Belgium administrative fiscal dataset and a household budget survey using a statistical matching procedure. That procedure enabled an imputation of personal tax liabilities to each household in the survey. Wagenhals (2004) compared several German
models. Each model attempted to combine some or all of the available income micro data; German Socio Economic Panel (GSOEP), Income and Consumption Survey (ICS) and Income Tax Panel (IAW). Lloyd (2007) elaborates on the use of a Static Income Model (STINMOD) developed by the National Centre for Social and Economic Modelling (NATSEM) in Australia. The base file was constructed from Australian Bureau of Statistics (ABS) surveys, namely the Survey of Incomes and Housing Costs (SIHC) and Household Expenditure Survey (HES). Lately, Eurostat of the European Commission (2013) issued guidelines for models based on data integration using statistical matching. It included a methodological overview and implementation of statistical matching and a case study on wage and labour statistics.

We applied important points learned from these previous studies in order handle the unique case of totally missing labour-income variables on the Susenas 2008. This was the start of our initial trial to build a better database for Indonesian income distribution studies and, in particular, to construct a 2008 PIT microsimulation base file using available Indonesian income micro data. Firstly, we prepared Susenas 2008 as our main base data for the analysis. Within this Susenas, we flagged observations with missing labour-income which were then matched with Sakernas 2008. Flagging was done at the individual level based on each individual working activity and the logical breakdown of the income detail. This was used for imputing labour-income for the completeness of personal income variables in the Susenas 2008. DGT administrative data 2008 of reported personal income tax return will augment the coverage of top income and reduce the survey data limitations. More specifically by exercising this imputation, top income earners from administrative data (who have an income higher than the highest income earner from the survey) will be added as new observations.
into the new dataset. This will develop a ‘true income’ approximation base-file for the analysis of Indonesian full income distribution.

2.3.1 Harmonisation and reconciliation of sources

Following D’Orazio, Di Zio & Scanu (2006), the first step for data matching is the harmonisation and reconciliation of the data sources. They highlight eight types of reconciliation starting with harmonisation of unit definition, reference period, target population, variables and classifications and then adjustment for accuracy, missing data and derivation of the variables.

In our case there is an important basic consideration which is the fact that both Sakernas and Susenas are household surveys conducted by BPS. Both have their own long history and have been improved over time. Hence we consider that Sakernas-Susenas 2008 matching is the best possible alternative for determining the true Indonesian income distribution for the given year and given available data for that year. We then need to further consider the main similarities and the main differences between the key concepts, definitions, and variables of the two sources (some of them have been previously detailed). Those similarities will be the supporting points for our matching exercise, while the differences will be handled by our harmonisation process to improve the comparability between the micro datasets.
Target Population

The Sakernas micro dataset is based on the individual unit observation. The targeted respondents are all members of households who are aged 10 years and above. Sakernas questionnaires are based on the labour and working activity definitions of the ILO.

Meanwhile Susenas has two types of micro datasets: Susenas-core which is an individual based dataset and Susenas-module which contains household-based micro data. Hence, the first step to prepare the Susenas income-consumption module for the recipient database is to adjust the unit of observations to individuals from previous household units. For this we took the detailed information of income and expenditure from households in the Susenas income-module 2008 and imputed them to the Susenas core-2008 individual unit by using household identification numbers as the unique id. The household identification number is created from a combination of code for province, municipality, district, sub-district, urban/rural, household number and person sampling number.

Through this exercise we merged the individual income-expenditure information from the Susenas module with the individual detailed socio economic characteristics in the Susenas-core. However, only individual labour income (all of which is missing for the 2008 data in particular) could be attached to individuals with employment. All other types of income and expenditure variables could only be allocated to the head of household by this matching id exercise. We then proceeded to the identification of the reference population represented in each micro dataset.
Table 1 shows absolute differences from the reference individual population 10 years of age and above. We should note that Susenas is a socio-economic survey that covers the whole of the Indonesian population. That means Susenas originally sampled all age ranges in the survey. The total sample from Susenas’ in 2008 was 282,387 households and 1,142,675 individuals. The sample represents 57,566,756 Indonesian households and 228,018,900 individuals. However, that is not the case with Sakernas; its reference population is different from Susenas. Sakernas follows ILO labour concepts and variables by not including individuals below 10 years of age in the sample.
Table 1. Comparison of weighted frequency for individual reference population by province between Sakernas & Susenas 2008

<table>
<thead>
<tr>
<th>Province</th>
<th>SAKERNAS 2008 (population 10 years of age and above)</th>
<th>SUSENAS 2008 (population 10 years of age and above)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no of observations (sample)</td>
<td>(weighted) absolute frequency</td>
<td>no of observations (sample)</td>
</tr>
<tr>
<td>NAD</td>
<td>39,097</td>
<td>3,422,783</td>
<td>38,538</td>
</tr>
<tr>
<td>Sumatera Utara</td>
<td>58,489</td>
<td>10,330,394</td>
<td>57,904</td>
</tr>
<tr>
<td>Sumatera Barat</td>
<td>36,585</td>
<td>3,824,057</td>
<td>36,306</td>
</tr>
<tr>
<td>Riau</td>
<td>23,089</td>
<td>4,072,047</td>
<td>22,934</td>
</tr>
<tr>
<td>Jambi</td>
<td>19,601</td>
<td>2,239,841</td>
<td>19,414</td>
</tr>
<tr>
<td>Sumatera Selatan</td>
<td>30,265</td>
<td>5,702,224</td>
<td>29,726</td>
</tr>
<tr>
<td>Bengkulu</td>
<td>19,619</td>
<td>1,322,671</td>
<td>17,561</td>
</tr>
<tr>
<td>Lampung</td>
<td>11,884</td>
<td>916,867</td>
<td>11,614</td>
</tr>
<tr>
<td>Bangka Belitung</td>
<td>11,810</td>
<td>1,144,973</td>
<td>11,556</td>
</tr>
<tr>
<td>Jakarta</td>
<td>21,942</td>
<td>7,690,258</td>
<td>22,834</td>
</tr>
<tr>
<td>Jawa Barat</td>
<td>64,559</td>
<td>33,403,900</td>
<td>64,446</td>
</tr>
<tr>
<td>Jawa Tengah</td>
<td>77,906</td>
<td>27,137,087</td>
<td>76,770</td>
</tr>
<tr>
<td>Yogyakarta</td>
<td>9,865</td>
<td>3,058,178</td>
<td>9,493</td>
</tr>
<tr>
<td>Jawa Timur</td>
<td>92,147</td>
<td>31,910,735</td>
<td>90,428</td>
</tr>
<tr>
<td>Banten</td>
<td>16,523</td>
<td>7,667,836</td>
<td>16,208</td>
</tr>
<tr>
<td>Bali</td>
<td>18,594</td>
<td>2,969,807</td>
<td>18,263</td>
</tr>
<tr>
<td>NTB</td>
<td>17,774</td>
<td>3,448,882</td>
<td>17,195</td>
</tr>
<tr>
<td>NTT</td>
<td>37,892</td>
<td>3,550,386</td>
<td>37,345</td>
</tr>
<tr>
<td>Kalimantan Barat</td>
<td>27,383</td>
<td>3,375,907</td>
<td>26,315</td>
</tr>
<tr>
<td>Kalimantan Tengah</td>
<td>26,191</td>
<td>1,651,376</td>
<td>26,047</td>
</tr>
<tr>
<td>Kalimantan Selatan</td>
<td>24,101</td>
<td>2,821,574</td>
<td>23,528</td>
</tr>
<tr>
<td>Kalimantan Timur</td>
<td>30,268</td>
<td>2,508,625</td>
<td>24,682</td>
</tr>
<tr>
<td>Sulawesi Utara</td>
<td>22,779</td>
<td>1,859,347</td>
<td>22,752</td>
</tr>
<tr>
<td>Sulawesi Tengah</td>
<td>20,533</td>
<td>1,968,522</td>
<td>19,868</td>
</tr>
<tr>
<td>Sulawesi Selatan</td>
<td>51,585</td>
<td>6,324,361</td>
<td>49,564</td>
</tr>
<tr>
<td>Sulawesi Tenggara</td>
<td>26,271</td>
<td>1,617,498</td>
<td>25,608</td>
</tr>
<tr>
<td>Gorontalo</td>
<td>12,314</td>
<td>788,164</td>
<td>12,598</td>
</tr>
<tr>
<td>Sulawesi Barat</td>
<td>10,660</td>
<td>837,229</td>
<td>10,344</td>
</tr>
<tr>
<td>Maluku</td>
<td>12,641</td>
<td>1,040,711</td>
<td>11,673</td>
</tr>
<tr>
<td>Maluku Utara</td>
<td>11,989</td>
<td>751,417</td>
<td>12,081</td>
</tr>
<tr>
<td>Papua Barat</td>
<td>7,619</td>
<td>575,984</td>
<td>6,947</td>
</tr>
<tr>
<td>Papua</td>
<td>17,389</td>
<td>1,636,279</td>
<td>16,241</td>
</tr>
<tr>
<td>TOTAL</td>
<td>931,890</td>
<td>187,553,465</td>
<td>909,149</td>
</tr>
</tbody>
</table>

Remark:
*) absolut frequency is the weighted number individual population
**) relative frequency is the proportion of the province population to the national population
Source: own simulation from Sakernas and Susenas 2008 micro dataset
In Table 1, we harmonise this difference by taking out the population below 10 years of age from our Susenas matching base. The comparison shows that, generally, the differences between the reference populations is not large, except for four provinces. The difference between the weighted frequencies of Bangka Belitung, Nusa Tenggara Timur, Kalimantan Tengah and Gorontalo province ranged from -12.32% to 11.63%. This difference might be related to the different sampling methods. Susenas uses multi stage-stratified sampling with municipality as the first stage sampling frame whereas Sakernas uses a two-stage sampling method with the master file by village as the first stage sampling frame. Further, according to BPS data, the Indonesian administrative areas as of July/August 2008 (in line with the Sakernas and Susenas periods) consisted of 495 municipalities/regencies and 6,579 sub-districts. Neither Sakernas nor Susenas completely covered all areas. Sakernas covered a total of 483 municipalities and regencies, which is better than Susenas with only 456. By contrast, in terms of sub-district coverage, Susenas did better than Sakernas. Susenas covered 5,419 sub-districts which was more than Sakernas with only 5,276.

In summary, between Sakernas and Susenas 2008, the total population differs by only 1.63%. Hence, we could confirm that we have a ‘similar’ reference population in both the donor dataset (Sakernas) and the recipient dataset (Susenas).

**Comparison of concept and distribution of common variables**

Following the finalisation of the individual-based Susenas dataset, we proceed to analyse relevant matching variables for the labour income imputation exercise. First we considered...
the relevant variables for tax microsimulation as our matching variables. We conducted the analysis for relevant matching variables by comparing the concept and distribution of common variables from Susenas and Sakernas. From the identified common variables, our choice of matching variables was socio-demographic variables which existed in both micro data sets and which used similar concepts to define them. The choices of matching variables are as follows:

- Geographic: province, municipality/regency, sub-district, village
- Demographic: highest education level attained, sex, age, marital status, number of household members
- Employment: active working status, main occupation sectors

We have confirmed the concept and definition, distribution and coverage of all geographic variables in the previous section. We further compare concept and definition for each of the similar variables between the two databases. We based our analysis on the BPS manuals for Sakernas 2008 and Susenas 2008. We generally found exactly the same definitions for most demographic and employment characteristics’ variables.

Minor exceptions were found for basic questions and the sub-coding of highest education level and on sub-coding of main occupation sectors. Susenas-core questionnaires asked for both ongoing highest education level and finished or attained highest education level, while Sakernas only asked about the level attained. Sub coding in Susenas differentiates between general and Islamic types of primary and secondary school, while Sakernas combines both categories. Sub coding of main occupation sectors only differs in one particular type of main occupation. Susenas has one additional category of other sectors which Sakernas did not
have. Yet, both Susenas and Sakernas have similar categories of unclassified sectors. In order to maintain consistency and comparability we created one new classification, ‘highest education level’, which accommodated both Susenas and Sakernas ‘attained highest education level’. We also created one other new sub-coding to combine Susenas unclassified and other sectors of main occupation. Our process of harmonising different basic questions and sub-coding of the highest education level variable and the main occupation sectors variable is shown in Table 2:
Table 2. Difference and Harmonisation of Variables’ Conceptual Differences and their Sub-Coding (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>SAKERNAS 2008</th>
<th>SUSENAS 2008</th>
<th>SAKERNAS &amp; SUSENAS 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highest Education Level (Attained)</strong></td>
<td><strong>Highest Education Level (Attained/Currently Ongoing)</strong></td>
<td><strong>New Classification (Highest Education Level)</strong></td>
</tr>
<tr>
<td>1 Not going / Never go to school</td>
<td>0 Not going / Never go to school</td>
<td>1 Not going / Never go to school &amp; Not finish/Not yet finished Primary School</td>
</tr>
<tr>
<td>2 Not finish / Not yet finish</td>
<td>Primary School</td>
<td>2 Primary School / Islamic Primary</td>
</tr>
<tr>
<td>3 Primary School / Islamic Primary</td>
<td>1 Primary School</td>
<td>2 Primary School / Islamic Primary</td>
</tr>
<tr>
<td>4 Junior High School / Islamic Junior High</td>
<td>3 Junior High School</td>
<td>3 Junior High School / Islamic Junior High</td>
</tr>
<tr>
<td>5 Vocational Junior High School</td>
<td>5 Vocational Junior High School</td>
<td>4 Vocational Junior High School</td>
</tr>
<tr>
<td>6 Senior High School / Islamic Senior High</td>
<td>6 Senior High School</td>
<td>5 Senior High School / Islamic Senior High</td>
</tr>
<tr>
<td>7 Vocational Senior High School</td>
<td>8 Vocational Senior High School</td>
<td>6 Vocational Senior High School</td>
</tr>
<tr>
<td>8 Diploma I / Diploma II</td>
<td>9 Diploma I / Diploma II</td>
<td>7 Diploma I / Diploma II</td>
</tr>
<tr>
<td>9 Diploma III</td>
<td>10 Diploma III</td>
<td>8 Diploma III</td>
</tr>
<tr>
<td>10 Diploma IV / Undergraduates</td>
<td>11 Diploma IV / Undergraduates</td>
<td>9 Diploma IV / Undergraduates</td>
</tr>
<tr>
<td>11 Post Graduates (Masters/Doctoral)</td>
<td>12 Post Graduates (Masters/Doctoral)</td>
<td>10 Post Graduates (Masters/Doctoral)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Main Occupation Sectors</th>
<th>Main Occupation Sectors</th>
<th>New Classification (Main Occupation Sectors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Unclassified</td>
<td>0 Unclassified</td>
<td>0 Unclassified</td>
</tr>
<tr>
<td>1 Agriculture, Plantation, Forest, Hunting &amp; Fisheries</td>
<td>1 Agriculture, Plantation, Forest, Hunting &amp; Fisheries</td>
<td>1 Agriculture, Plantation, Forest, Hunting &amp; Fisheries</td>
</tr>
<tr>
<td>2 Quarrying &amp; Mining</td>
<td>2 Quarrying &amp; Mining</td>
<td>2 Quarrying &amp; Mining</td>
</tr>
<tr>
<td>3 Manufacturing Industry</td>
<td>3 Manufacturing Industry</td>
<td>3 Manufacturing Industry</td>
</tr>
<tr>
<td>4 Electricity, Gas, Water</td>
<td>4 Electricity, Gas, Water</td>
<td>4 Electricity, Gas, Water</td>
</tr>
<tr>
<td>5 Construction</td>
<td>5 Construction</td>
<td>5 Construction</td>
</tr>
<tr>
<td>6 Trade, Restaurant &amp; Accomodation Services</td>
<td>6 Trade, Restaurant &amp; Accomodation Services</td>
<td>6 Trade, Restaurant &amp; Accomodation Services</td>
</tr>
<tr>
<td>7 Transportation, Warehouse &amp; Communication</td>
<td>7 Transportation, Warehouse &amp; Communication</td>
<td>7 Transportation, Warehouse &amp; Communication</td>
</tr>
<tr>
<td>8 Financial Services, Real Estate, Rental &amp; Services</td>
<td>8 Financial Services, Real Estate, Rental &amp; Services</td>
<td>8 Financial Services, Real Estate, Rental &amp; Services</td>
</tr>
<tr>
<td>9 Social, Community &amp; Personal Services</td>
<td>9 Social, Community &amp; Personal Services</td>
<td>9 Social, Community &amp; Personal Services</td>
</tr>
</tbody>
</table>

Secondly we compared the relative and absolute frequency distributions between the matching variables. These results are shown in Table 3.
Demographic: highest educational level attained

Table 3. Comparison of weighted frequency for individual reference population after harmonisation of highest educational level variable (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>Highest Education Level</th>
<th>SAKERNAS 2008 (population 10 years of age and above)</th>
<th>SUSenas 2008 (population 10 years of age and above)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no of observations (sample)</td>
<td>(weighted) absolute frequency</td>
<td>no of observations (sample)</td>
</tr>
<tr>
<td>Not going/never go to school &amp; Not finish/still in Primary School</td>
<td>222,350</td>
<td>40,789,494</td>
<td>264,889</td>
</tr>
<tr>
<td>Primary School/Islamic Primary</td>
<td>318,701</td>
<td>65,181,997</td>
<td>265,161</td>
</tr>
<tr>
<td>Junior High School/Islamic Junior High</td>
<td>168,534</td>
<td>35,556,578</td>
<td>157,559</td>
</tr>
<tr>
<td>Vocational Junior High School</td>
<td>8,893</td>
<td>1,687,757</td>
<td>6,605</td>
</tr>
<tr>
<td>Senior High School/Islamic Senior High</td>
<td>122,420</td>
<td>24,364,803</td>
<td>131,639</td>
</tr>
<tr>
<td>Vocational Senior High School</td>
<td>46,990</td>
<td>10,725,621</td>
<td>41,053</td>
</tr>
<tr>
<td>Diploma I / Diploma II</td>
<td>9,595</td>
<td>1,687,553</td>
<td>8,003</td>
</tr>
<tr>
<td>Diploma III</td>
<td>9,599</td>
<td>2,184,221</td>
<td>8,688</td>
</tr>
<tr>
<td>Diploma IV/Undergraduates</td>
<td>23,293</td>
<td>5,021,418</td>
<td>24,261</td>
</tr>
<tr>
<td>Post Graduates (Masters/Doctoral)</td>
<td>1,515</td>
<td>354,023</td>
<td>1,291</td>
</tr>
<tr>
<td>TOTAL</td>
<td>931,890</td>
<td>187,553,465</td>
<td>909,149</td>
</tr>
</tbody>
</table>

The most notable difference was for individuals in the first category who have not finished primary school. In this category, Susenas sampled 25 percent more individuals than Sakernas did. The next big difference was individuals with 1 year and/or 2 year diploma education. In this case Sakernas covered 18 percent more individuals than Susenas. The other big difference, which ranged between 12% and 14%, was in the sub category of individuals who only finished primary school, junior high school, junior vocational high school and senior vocational high school. In general, we concluded that Sakernas covered more individuals with higher education except for the undergraduate category.
The fact that the difference in the distributions between the two surveys is relatively high does not mean this variable is omitted from the list of matching variables. This variable is one of the important determinants for an individual’s income level and we might expect a better matching result from our harmonisation effort after including the highest education level variable. To ensure that this variable does assist the matching process, we will examine the result of the statistical matching by disaggregating the highest educational level category.

**Demographic: age group and sex**

This category follows the national publication on population statistics. BPS disaggregates population based on age group and sex. For this reason we compared the distribution of population by age group and sex between our two data sets. We could argue that these age group and sex variables should be compared with the distribution from the population census, however there is little point in doing this because the master survey frame of the Indonesian survey itself came from the census and the result has been benchmarked to the census populations by the BPS.

Looking at table 4, we can see a significant difference between the male and female distribution, especially in the oldest age group which is the above 75 years old age group. Most of the young age group show a relatively small difference except for the category of 15 – 20 years which has a 10.5% difference in the male distribution between the two. Fifteen years old is considered the starting age for entry into the labour force if not studying (a category of an active labour population). We later analysed how this difference between the two distributions affected the statistical matching of the active labour population. In sum, we
confirmed that the distribution of males and females between different age groups was similar from the micro dataset.

Table 4. Comparison of weighted frequency for individual reference population by age group and sex variable (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>Age Group</th>
<th>(Weighted) Absolute Frequency</th>
<th>% Difference</th>
<th>(Weighted) Absolute Frequency</th>
<th>% Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-15</td>
<td>SAKERNAS 2008: 24,619,368</td>
<td>2.72%</td>
<td>SUSENAS 2008: 23,949,928</td>
<td>2.79%</td>
</tr>
<tr>
<td></td>
<td>15-20 SAKERNAS 2008: 9,875,787</td>
<td>10.54%</td>
<td>SUSENAS 2008: 9,096,723</td>
<td>8.14%</td>
</tr>
<tr>
<td></td>
<td>20-25 SAKERNAS 2008: 10,075,783</td>
<td>9.43%</td>
<td>SUSENAS 2008: 9,610,887</td>
<td>8.07%</td>
</tr>
<tr>
<td></td>
<td>25-30 SAKERNAS 2008: 9,415,908</td>
<td>9.02%</td>
<td>SUSENAS 2008: 9,740,504</td>
<td>6.38%</td>
</tr>
<tr>
<td></td>
<td>30-35 SAKERNAS 2008: 8,656,885</td>
<td>-0.22%</td>
<td>SUSENAS 2008: 8,949,668</td>
<td>0.20%</td>
</tr>
<tr>
<td></td>
<td>35-40 SAKERNAS 2008: 7,562,197</td>
<td>0.08%</td>
<td>SUSENAS 2008: 7,746,568</td>
<td>-2.07%</td>
</tr>
<tr>
<td></td>
<td>40-45 SAKERNAS 2008: 6,417,980</td>
<td>-4.24%</td>
<td>SUSENAS 2008: 6,810,519</td>
<td>-5.94%</td>
</tr>
<tr>
<td></td>
<td>45-50 SAKERNAS 2008: 5,444,569</td>
<td>-1.98%</td>
<td>SUSENAS 2008: 5,299,609</td>
<td>-8.13%</td>
</tr>
<tr>
<td></td>
<td>50-55 SAKERNAS 2008: 3,856,573</td>
<td>-7.28%</td>
<td>SUSENAS 2008: 4,021,155</td>
<td>-7.29%</td>
</tr>
<tr>
<td></td>
<td>55-60 SAKERNAS 2008: 2,758,605</td>
<td>-4.30%</td>
<td>SUSENAS 2008: 2,908,760</td>
<td>-9.98%</td>
</tr>
<tr>
<td></td>
<td>60-65 SAKERNAS 2008: 2,052,249</td>
<td>-9.47%</td>
<td>SUSENAS 2008: 2,534,451</td>
<td>-12.23%</td>
</tr>
<tr>
<td></td>
<td>65-70 SAKERNAS 2008: 1,312,892</td>
<td>-11.61%</td>
<td>SUSENAS 2008: 1,732,072</td>
<td>-6.22%</td>
</tr>
<tr>
<td></td>
<td>70-75 SAKERNAS 2008: 810,208</td>
<td>-9.43%</td>
<td>SUSENAS 2008: 1,045,615</td>
<td>-11.13%</td>
</tr>
<tr>
<td></td>
<td>75-80 SAKERNAS 2008: 381,100</td>
<td>-13.17%</td>
<td>SUSENAS 2008: 536,546</td>
<td>-21.62%</td>
</tr>
<tr>
<td></td>
<td>90-95 SAKERNAS 2008: 33,270</td>
<td>-32.93%</td>
<td>SUSENAS 2008: 67,316</td>
<td>-7.29%</td>
</tr>
<tr>
<td>&gt;95</td>
<td>SAKERNAS 2008: 36,822</td>
<td>-17.45%</td>
<td>SUSENAS 2008: 85,046</td>
<td>-34.63%</td>
</tr>
</tbody>
</table>

Note that age will also serve as a determinant for allocating the number of dependents to income earners and tax payers in our personal income tax microsimulation that was built on our imputation datasets. Later, in Chapter III, there is further elaboration on this.
Demographic: marital status

Marital status is another important variable used in the income tax microsimulation process. Marital status determines the entitlement to non-taxable income for the individual tax payer.

Table 5. Comparison of weighted frequency for individual reference population by marital status variable (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>Marital Status</th>
<th>SAKERNAS 2008 (population 10 years of age and above)</th>
<th>SUSENAS 2008 (population 10 years of age and above)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no of observations (sample)</td>
<td>(weighted) absolute frequency</td>
<td>no of observations (sample)</td>
</tr>
<tr>
<td>Not Married</td>
<td>325,668</td>
<td>64,722,939</td>
<td>318,328</td>
</tr>
<tr>
<td>Married</td>
<td>538,052</td>
<td>109,206,148</td>
<td>524,354</td>
</tr>
<tr>
<td>Divorced</td>
<td>14,103</td>
<td>2,990,561</td>
<td>14,645</td>
</tr>
<tr>
<td>Widow (death of the spouse)</td>
<td>54,067</td>
<td>10,633,817</td>
<td>51,822</td>
</tr>
<tr>
<td>TOTAL</td>
<td>187,553,465</td>
<td>37,508,395</td>
<td>1.62%</td>
</tr>
</tbody>
</table>

The insignificant differences between the two data sets presented in table 5 confirms that the Sakernas and Susenas distributions of marital status are generally similar. Hence we also confirm the marital status variable as one of our matching variables.

Demographic: number of household members

This variable, the number of household members is also a determinant in the later PIT microsimulation. Table 6 shows that there are small differences between the two datasets.
Table 6. Comparison of weighted frequency for individual reference population by number of household member (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>No of Household Member</th>
<th>SAKERNAS 2008 (population 10 years of age and above)</th>
<th>SUSENAS 2008 (population 10 years of age and above)</th>
<th>Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no of observations (sample)</td>
<td>(weighted) absolute frequency</td>
<td>no of observations (sample)</td>
</tr>
<tr>
<td>1</td>
<td>17,868</td>
<td>3,840,255</td>
<td>15,832</td>
</tr>
<tr>
<td>2</td>
<td>74,606</td>
<td>15,909,745</td>
<td>70,258</td>
</tr>
<tr>
<td>3</td>
<td>158,367</td>
<td>34,917,317</td>
<td>154,222</td>
</tr>
<tr>
<td>4</td>
<td>222,086</td>
<td>47,523,294</td>
<td>215,469</td>
</tr>
<tr>
<td>5</td>
<td>189,649</td>
<td>38,003,087</td>
<td>184,843</td>
</tr>
<tr>
<td>6</td>
<td>124,930</td>
<td>23,225,624</td>
<td>124,829</td>
</tr>
<tr>
<td>7</td>
<td>70,523</td>
<td>12,181,406</td>
<td>70,567</td>
</tr>
<tr>
<td>8</td>
<td>37,095</td>
<td>6,027,316</td>
<td>37,260</td>
</tr>
<tr>
<td>9</td>
<td>17,838</td>
<td>2,956,860</td>
<td>18,327</td>
</tr>
<tr>
<td>&gt;9</td>
<td>18,928</td>
<td>2,968,561</td>
<td>17,542</td>
</tr>
<tr>
<td>TOTAL</td>
<td>187,553,465</td>
<td>184,508,395</td>
<td>1.62%</td>
</tr>
</tbody>
</table>

Because of these small differences we used the distribution of the number of household members as one of our matching variables.

This also enabled us to proceed to the important step of developing the income tax base which will be discussed in chapter II, to find the number of income earners in the household and the number of dependents. This number of dependents will be allocated to individuals who earned income in the household.

**Employment: active working status**

Working status is one of the major indicators for our labour income imputation. In particular, it was used used in flagging the Susenas individual observation for labour income-imputation
from the Sakernas donor dataset. The data in table 7 show that Sakernas coverage of working population is less than for Susenas by 3.78%. To further examine the difference we analysed the detailed breakdown of the active working population. The breakdown of labour force classification is based on the Sakernas concept, which is an adoption of the ILO concept of labour variables. Accordingly, the labour force is defined as the population who are 15 years of age and above, who actively work (35 hours per week) or have a job but are temporarily not working. After the bottom range was set, we then set 60 years as a maximum age for the working population. This is based on the average of the Indonesian retirement age, and follows the Ministry of Manpower (for private sector) and Ministry of State Apparatus regulations (for public servant).

Table 7. Comparison of weighted frequency for individual reference population by active working status (Sakernas & Susenas 2008)

| Working Status | SAKERNAS 2008 (population 10 years of age and above) | SUSENAS 2008 (population 10 years of age and above) | Difference  
|----------------|-----------------------------------------------------|-----------------------------------------------------|------------
|                | no of observations (sample) | (weighted) absolute frequency | no of observations (sample) | (weighted) absolute frequency | (weighted) absolute frequency |
| Working        | 507,212 | 101,214,521 | 525,954 | 105,045,036 | -3.78% |
| Not Working    | 424,678 | 86,338,944 | 383,195 | 79,463,359 | 7.96% |
| TOTAL          | 187,553,465 | 184,508,395 | 1.62% |

We also found that the population from 15 years old up to 60 years old accounts for 87.7% of the Sakernas working population and 86.8% of the Susenas working population. Since the coverage of the Sakernas active working population is less than the Susenas active working population, the two datasets differ (by 2.64%) in favour of Susenas. This explains the major proportion of the difference in the active working population.
Employment: main occupation sectors

From 2005 until 2011, Sakernas was done each semester in February and August with a sample of around 67,000 while Susenas 2008 was done in July with a sample of 26,000. In this thesis we did not investigate any impact of seasonality on occupation sectors. Indonesia is less affected by seasonality than other countries due to its tropical location. Issues may arise if the survey was being done at the time of national festivity such as independence day or religious festivals.

Some significant differences can be seen in five out of nine sub categories of the main occupation sectors (Table 8). We see big differences in Sakernas, especially for respondents who work in several highly-paid sectors like mining/quarrying (-22.96%) and financial sectors (-34.74%). The biggest difference is noted in electricity, gas and water employees where Susenas estimated almost twice the respondents of Sakernas. This may have an impact on the matching result after disaggregating by main occupation sector and at the end of our matching process we considered how to proceed with this. At this point we decided to use this variable as one of our matching variables, as we believe that failure to consider this major employment variable would significantly compromise the quality of our matching result. Later in table 15 we will look at the difference in incomes for different occupations from the matching result.

Table 8. Comparison of weighted frequency for individual reference population by main
occupation sectors (Sakernas & Susenas 2008)

<table>
<thead>
<tr>
<th>Main Occupation Sectors</th>
<th>SAKERNAS 2008 (population 10 years of age and above)</th>
<th>SUSENAS 2008 (population 10 years of age and above)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no of observations (sample)</td>
<td>(weighted) absolute frequency</td>
<td>no of observations (sample)</td>
</tr>
<tr>
<td>Unclassified</td>
<td>413,198</td>
<td>84,102,780</td>
<td>376,940</td>
</tr>
<tr>
<td>Agriculture, Plantation, Forest, Hunting &amp; Fisheries</td>
<td>252,628</td>
<td>41,866,557</td>
<td>261,068</td>
</tr>
<tr>
<td>Quarrying &amp; Mining</td>
<td>7,653</td>
<td>1,080,092</td>
<td>9,203</td>
</tr>
<tr>
<td>Manufacturing Industry</td>
<td>44,785</td>
<td>12,660,812</td>
<td>41,326</td>
</tr>
<tr>
<td>Electricity, Gas, Water</td>
<td>1,066</td>
<td>201,172</td>
<td>1,817</td>
</tr>
<tr>
<td>Construction</td>
<td>24,006</td>
<td>5,446,786</td>
<td>21,390</td>
</tr>
<tr>
<td>Trade, Restaurant &amp; Accomodation Services</td>
<td>91,055</td>
<td>21,404,723</td>
<td>88,616</td>
</tr>
<tr>
<td>Transportation, Warehouse &amp; Communication</td>
<td>27,790</td>
<td>6,187,321</td>
<td>21,748</td>
</tr>
<tr>
<td>Financial Services, Real Estate, Rental &amp; Services</td>
<td>5,369</td>
<td>1,461,233</td>
<td>7,112</td>
</tr>
<tr>
<td>Social, Community &amp; Personal Services</td>
<td>64,340</td>
<td>13,141,989</td>
<td>79,929</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>187,553,465</td>
<td>184,508,395</td>
<td></td>
</tr>
</tbody>
</table>

**Wage concept and reference period**

The concept of current salary/wages from employment is different for the two micro data sources. Sakernas recorded net income from employment, while Susenas recorded gross income from employment. Referring to the Sakernas manual, labour or employment income is salary/wages in the form of money and/or goods from the main job which is recorded net after tax and after other deductions. Examples of other deductions are compulsory employment insurance and pension premiums. On the other hand, Susenas labour income is salary/wages in the form of money and/or goods/services from both main and secondary jobs (recorded separately). Susenas labour income is recorded gross before tax and other employment related deductions. The main difference is in the gross income, income from
secondary jobs and inclusion of salary in the forms of services. Salary in the form of services could be explained with an example of a company that provides housing facilities. If this facility is given by the company then the imputed monthly rent for this is calculated and included in the component of income from employment. The reference periods for both Sakernas and Susenas labour income are the same (i.e. last monthly salary/wage received). The harmonisation process was needed to overcome this conceptual problem, and this is explained in the next section on income grossing-up.

2.3.2 Income Alignment and Grossing Up

Official BPS average labour wage per province is published annually. We used these national and publicly available data as the benchmark for aligning our Sakernas micro data.

Table 9. Difference between national publications on average labour wages per province vs weighted average labour wages per province (SAKERNAS 2008 Micro Data)

<table>
<thead>
<tr>
<th>Code</th>
<th>Province</th>
<th>Average Labour Salary</th>
<th>Weighted Average Labour Wage</th>
<th>Diff</th>
<th>Code</th>
<th>Province</th>
<th>Average Labour Salary</th>
<th>Weighted Average Labour Wage</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>NAD</td>
<td>1,361,664</td>
<td>1,174,534</td>
<td>13.74%</td>
<td>52</td>
<td>NTB</td>
<td>1,060,457</td>
<td>691,014</td>
<td>34.84%</td>
</tr>
<tr>
<td>12</td>
<td>Sumatera Utara</td>
<td>1,107,912</td>
<td>983,127</td>
<td>11.26%</td>
<td>53</td>
<td>NTT</td>
<td>1,229,845</td>
<td>847,097</td>
<td>31.12%</td>
</tr>
<tr>
<td>13</td>
<td>Sumatera Barat</td>
<td>1,213,302</td>
<td>1,024,867</td>
<td>15.53%</td>
<td>61</td>
<td>Kalimantan Barat</td>
<td>1,178,395</td>
<td>1,086,154</td>
<td>7.83%</td>
</tr>
<tr>
<td>14</td>
<td>Riau</td>
<td>1,345,689</td>
<td>1,273,020</td>
<td>5.40%</td>
<td>62</td>
<td>Kalimantan Tengah</td>
<td>1,127,597</td>
<td>1,138,004</td>
<td>-0.92%</td>
</tr>
<tr>
<td>15</td>
<td>Jambi</td>
<td>1,047,964</td>
<td>1,088,860</td>
<td>-3.90%</td>
<td>63</td>
<td>Kalimantan Selatan</td>
<td>1,134,197</td>
<td>1,005,636</td>
<td>11.33%</td>
</tr>
<tr>
<td>16</td>
<td>Sumatera Selatan</td>
<td>1,107,029</td>
<td>979,986</td>
<td>11.48%</td>
<td>64</td>
<td>Kalimantan Timur</td>
<td>1,806,191</td>
<td>1,602,739</td>
<td>11.26%</td>
</tr>
<tr>
<td>17</td>
<td>Bengkulu</td>
<td>1,291,764</td>
<td>1,034,181</td>
<td>19.94%</td>
<td>65</td>
<td>Sulawesi Utara</td>
<td>1,220,020</td>
<td>1,009,680</td>
<td>17.24%</td>
</tr>
<tr>
<td>18</td>
<td>Lampung</td>
<td>967,877</td>
<td>773,360</td>
<td>20.10%</td>
<td>66</td>
<td>Sulawesi Tengah</td>
<td>1,135,869</td>
<td>900,419</td>
<td>20.73%</td>
</tr>
<tr>
<td>19</td>
<td>Bangka Belitung</td>
<td>1,071,139</td>
<td>1,161,000</td>
<td>-8.39%</td>
<td>71</td>
<td>Sulawesi Selatan</td>
<td>1,184,837</td>
<td>990,650</td>
<td>16.39%</td>
</tr>
<tr>
<td>21</td>
<td>Kep Riau</td>
<td>1,643,156</td>
<td>1,616,573</td>
<td>1.62%</td>
<td>72</td>
<td>Sulawesi Tenggara</td>
<td>1,182,569</td>
<td>956,837</td>
<td>19.09%</td>
</tr>
<tr>
<td>31</td>
<td>Jakarta</td>
<td>1,632,435</td>
<td>1,578,140</td>
<td>3.33%</td>
<td>75</td>
<td>Gorontalo</td>
<td>911,867</td>
<td>776,583</td>
<td>14.84%</td>
</tr>
<tr>
<td>32</td>
<td>Jawa Barat</td>
<td>1,190,466</td>
<td>931,290</td>
<td>21.77%</td>
<td>76</td>
<td>Sulawesi Barat</td>
<td>1,166,340</td>
<td>771,270</td>
<td>33.87%</td>
</tr>
<tr>
<td>33</td>
<td>Jawa Tengah</td>
<td>848,534</td>
<td>688,290</td>
<td>18.88%</td>
<td>81</td>
<td>Maluku</td>
<td>1,326,904</td>
<td>980,599</td>
<td>26.10%</td>
</tr>
<tr>
<td>34</td>
<td>Yogyakarta</td>
<td>1,061,581</td>
<td>898,236</td>
<td>15.39%</td>
<td>82</td>
<td>Maluku Utara</td>
<td>1,310,048</td>
<td>1,144,428</td>
<td>12.64%</td>
</tr>
<tr>
<td>35</td>
<td>Jawa Timur</td>
<td>930,142</td>
<td>731,352</td>
<td>21.37%</td>
<td>91</td>
<td>Papua Barat</td>
<td>1,689,114</td>
<td>1,457,094</td>
<td>13.74%</td>
</tr>
<tr>
<td>36</td>
<td>Banten</td>
<td>1,275,265</td>
<td>1,113,023</td>
<td>12.72%</td>
<td>94</td>
<td>Papua</td>
<td>2,102,546</td>
<td>1,635,072</td>
<td>22.23%</td>
</tr>
<tr>
<td>51</td>
<td>Bali</td>
<td>1,248,054</td>
<td>1,045,754</td>
<td>16.21%</td>
<td></td>
<td>Total National</td>
<td>1,158,085</td>
<td>947,000</td>
<td>18.23%</td>
</tr>
</tbody>
</table>

(*) Source: Central Bureau of Statistics (BPS), the Republic of Indonesia, Trend of Selected Socio-Economic Indicators of Indonesia (2009). Indonesia: Author
From Table 9 we can see various significant differences between provincial average labour data from the national BPS publication and the micro data source provided by Sakernas 2008. This means that we needed to align our Sakernas micro data to match the national benchmark. We aligned all individual observations by multiplying each individual labour salary by the provincial up-rate factor. This up-rate factor was calculated using the following formula:

\[
\text{Provincial up-rate factor} = 1 + \left(\text{percentage difference between weighted average labour wage Sakernas 2008 and Average Provincial Labour Salary 2008 from BPS national publication}\right)
\]

**Grossing Up**

After the income was aligned, we needed to consider that the reported income in the Sakernas is net income and take-home pay of the respondent which is different to the Susenas reported gross-income. We needed to harmonise the difference in order to impute the Susenas missing labour income information to produce a better estimate of Indonesian full income distribution.

Based on the available information in Sakernas, we needed to gross up our aligned income by assuming that the only reduction of salary/wages resulted from tax relief in the form of non-taxable income entitlements. The other gross up components which should be taken into consideration are occupational, pension cost, compulsory social insurance. This occupational and pension cost is a gross income deduction which is required by law. It comprises 5 percent of the gross income or gross pension payment with a different maximum cap for each. Other deductions could also come from the pension contribution paid by the employee. Occupational and pension cost’s apply evenly to both private and public employees while
deductions for compulsory social insurance consist of accident, death, health insurance paid by the employer, and pension fund paid by the employer. These deductions for compulsory social insurance are components of employee gross income. These applied differently to private and public employees. Private sector social insurance is regulated in accordance with article 20 Law no 3(1992). The detail is elaborated in the Government regulation no 14(1993) article 9. It explains five levels of accident/safety insurance from 0.24% to 1.27% of monthly wage. It also explains pension insurance borne by the employer (3.7% of monthly wage) and employee (2% of monthly wage). Death insurance (0.3% of monthly wage) is paid by the employer and health insurance of 6% of monthly wage for married employees and 3% of monthly wage for single workers, both paid by the employer. Public sector social insurance as regulated with Presidential decree no 8(1977) and consists of 4.75% pension fund instalment, 3.25% instalment for old age and 2% for health funds. All of these gross up components only apply to full time employees and pensioners.

We omitted these components in our formula for grossing up net income information. There was a lack information on the full-time and non-full-time employees and the need to build additional assumptions for different rates applied for private and public sectors employee complicated the calculation. There was also a further complication due to the need to assume a pensioner age that could be be built into both our gross up estimator and our tax estimator. In order to minimize the impact of these assumptions we did not put this eligible deduction into our tax estimator (see Chapter III).

Determinants of non-taxable income entitlement in Sakernas data are: information in relation to head of household, sex, age, marital status, and number of household members. At the first
stage, we identified the income and non-income earners in the household using income tax law criteria. We then identified possible dependents from the non-income earners in the household. Those were all persons under 18 years of age or a person over 18 years old but earning no income and not yet married. These possible dependents were then allocated to each of the income earners in the household. For this allocation, each of the income earners in the household was entitled to have a maximum of three dependents (following the income tax rule). We based our allocation on hierarchical relationships starting with allocating the maximum number of dependents to the head of household as the first income earner. We then moved to the other income earners of the household to allocate the remaining possible dependents. As noted on the allocation of dependents, we limited the dependents for the working wife or for the woman as the income earner to only herself unless this income earner was also the head of household.

In line with previous assumptions, we also assumed no combined income reporting in the household. This means that we treated each income earner in the household as a separate taxpayer who filed their own tax return. We omitted the possibility of a combined income of husband and wife. Even if this does not entirely represent all the possible treatments allowed by the law, in the end the net aggregate result will be the same. Most importantly, both of the treatments are permitted by the law. We had to limit our simulations to avoid unnecessary additional assumptions which could complicate our base data.

After this final assumption was built into the model, we then developed our gross-up formula to calculate the gross income of each of the tax payers in the Sakernas data. Gross income was estimated by an iterative process using available information being net income, personal
exemptions, non-taxable income, and the marginal tax rate. We divided our gross-up process into two groups; the first one was gross-up income based on Law no 17(2000) and the second one was based on Law no 36(2008). The gross up equations also identify which individuals earn gross salaries above a taxable income threshold under the previous Law no 17(2000) and under the latest Law no 36(2008). The gross up procedure is shown in Figure 2.
Figure 2. SAKERNAS (2008) Grossing Up Process

- Identify salary and non-salary earners
- Identify possible dependent from non-salary earners
- Allocate possible dependent to each salary earner in the household
- Generate non-taxable income based on old law
- Generate non-taxable income based on new law
- Grossing-up income based on old law
- Grossing-up income based on new law
- Merge

### Income Bracket (Rp)

<table>
<thead>
<tr>
<th>Status</th>
<th>Dependent</th>
<th>Non Tax Inc (Rp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Married</td>
<td>0</td>
<td>13,200,000</td>
</tr>
<tr>
<td>Not Married</td>
<td>0</td>
<td>14,400,000</td>
</tr>
<tr>
<td>Married</td>
<td>1</td>
<td>15,600,000</td>
</tr>
<tr>
<td>Married</td>
<td>2</td>
<td>16,800,000</td>
</tr>
<tr>
<td>Married</td>
<td>3</td>
<td>18,000,000</td>
</tr>
</tbody>
</table>

### Status Dependent Non Tax Inc (Rp)

- Not Married, 0: 13,200,000
- Not Married, 0: 14,400,000
- Married, 1: 15,600,000
- Married, 2: 16,800,000
- Married, 3: 18,000,000

### Rate (%) Formula

- **Below threshold**: $\text{grossinc} = \text{netinc}$
- **0 - 25 million**: $\text{grossinc} = \frac{\text{netinc} - 5\% \times \text{nontaxinc}}{1 - 5\%}$
- **25 - 50 million**: $\text{grossinc} = \frac{\text{netinc} - 12,500,000 - 10\% \times \text{nontaxinc}}{1 - 10\%}$
- **50 - 100 million**: $\text{grossinc} = \frac{\text{netinc} - 37,500,000 - 15\% \times \text{nontaxinc}}{1 - 15\%}$
- **100 - 200 million**: $\text{grossinc} = \frac{\text{netinc} - 137,500,000 - 25\% \times \text{nontaxinc}}{1 - 25\%}$
- **200 million more**: $\text{grossinc} = \frac{\text{netinc} - 337,500,000 - 35\% \times \text{nontaxinc}}{1 - 35\%}$

### Grossing-up Income

- Grossing-up income based on old law
- Grossing-up income based on new law

### Merge

- SAKERNAS 2008 Aligned and Grossed Up

---

netinc = annual net income
nontaxinc = non taxable income

grossinc = gross income

- Rate (%)
- Formula
- Below threshold: $\text{grossinc} = \text{netinc}$
- 0 - 25 million: $\text{grossinc} = \frac{\text{netinc} - 5\% \times \text{nontaxinc}}{1 - 5\%}$
- 25 - 50 million: $\text{grossinc} = \frac{\text{netinc} - 12,500,000 - 10\% \times \text{nontaxinc}}{1 - 10\%}$
- 50 - 100 million: $\text{grossinc} = \frac{\text{netinc} - 37,500,000 - 15\% \times \text{nontaxinc}}{1 - 15\%}$
- 100 - 200 million: $\text{grossinc} = \frac{\text{netinc} - 137,500,000 - 25\% \times \text{nontaxinc}}{1 - 25\%}$
- 200 million more: $\text{grossinc} = \frac{\text{netinc} - 337,500,000 - 35\% \times \text{nontaxinc}}{1 - 35\%}$

netinc = annual net income
nontaxinc = non taxable income

grossinc = gross income
Following the construction of a new variable of gross-income under the previous and new law, each of the grossed-up results were then merged. At this stage, the base data was ready for the next step.

2.3.3 Imputation of Labour Income

The next step in building the base dataset was a matching process to impute the missing labour income information in Susenas 2008. Fortunately, the labour-income earner could be identified among the respondents. However, in some cases labour-income information was missing from the labour-income earner. As explained in our previous section, labour-income was not available on the Susenas 2008 survey. In order to conduct personal income tax microsimulation, we used the Sakernas 2008 as the donor to complete Susenas income information for the reference year.

We imputed information on wage data in Susenas 2008 by drawing on the income information from Sakernas 2008. Following a search of the literature (Bennet, 2001; Gavin, 1985; Nordholt, 1998; OECD, 2013; Radner, 1981) we chose to use a modified cold-deck matching method. Cold-deck is usually used for imputing values onto panel surveys whereby the values of the same person from another wave of surveys is used to impute the missing values in the survey for the reference year. Our cold-deck sorted the data according to the matching variables in the independent donor survey using information from the similar matched record and transferred the income variable to the nearest match in the primary
survey data. By modified cold-deck we mean that the income value of other (but similar) persons in the independent survey (Sakernas 2008) are imputed to Susenas 2008.

As explained by Nordholt (1998) we adopted the application of matching key variables or covariates from both Sakernas and Susenas 2008 in order to complete the records in Susenas 2008. Records were matched when they had the same values on the covariates. In our case, the covariates are geographic (the region where the respondent live) and demographic (characteristics of the respondent and their economic sectors of employment). There is a high probability that two records with the same scores on the covariates will have the same or almost equal values of income.
Modified Cold-deck Matching Method

Figure 3. Step 1 to 3 of the matching process.

1. Identify imputing variables and decide the sequence of imputing variables:
   - Geographical variable:
     - Province
     - Municipality
     - District
     - Sub District
   - Correlation test with income/salary of SAKERNAS 2008:
     - Work Activity (-0.3211)
     - Main Occupation Sectors (0.4315)
     - Highest Education Level (0.308)
     - Sex (-0.1713)
     - Age (0.0909)
     - Marital Stat (0.0664)
     - No of household Member (-0.0285)
     - Dependent (modified variables a function of no of household member adjusted with PIT rulings).

2. Generate combination of imputing variable and merge 1 to 1 sequentially using each combination of the variable below:
   - Province
   - Municipality
   - District
   - Sub District
   - Work Activity
   - Main Occupation Sectors
   - Highest Education Level
   - Sex
   - Age
   - Marital Stat
   - No of household Member
   - Dependent

3. Basedata Susenas 2008 (Individual observation unit)

   Susenas 2008 Unmatched
   Susenas 2008 Matched

   Basedata Sakernas 2008 (Individual observation unit)

   Sakernas 2008 Unmatched
Identification of imputing variables and the sequence of the variables.

We set up geographic variables as our first linking variables. Then, from the other identified demographic and employment variables, we selected a sequence of variables to use. This sequence determined the variables which were linked first and which were linked last in the data matching process. We also decided how to arrange the data matching sequence, between the master and donor file respectively. For that we required the most significant correlation coefficient between each of the Sakernas 2008 linking variables to total salary/wages. We used the standard formula for the calculation of the correlation coefficient for the survey data.

\[
\rho = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2 \sum_{i=1}^{N}(X_i - \bar{X})^2}}
\]

Our result showed the order of the variable starting with the highest correlation coefficient as follows:

<table>
<thead>
<tr>
<th>Demographic &amp; Employment Variables</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>main occupation sector</td>
<td>0.4315</td>
</tr>
<tr>
<td>activity work</td>
<td>-0.3211</td>
</tr>
<tr>
<td>highest education level</td>
<td>0.308</td>
</tr>
<tr>
<td>sex</td>
<td>-0.1713</td>
</tr>
<tr>
<td>age</td>
<td>0.0909</td>
</tr>
<tr>
<td>marital status</td>
<td>0.0664</td>
</tr>
<tr>
<td>no of household members</td>
<td>-0.0285</td>
</tr>
</tbody>
</table>

From there, we sorted our donor (Sakernas 2008) and our master file (Susenas 2008) based on geographic and imputing variables. We sorted our geographic variables based on province, municipality, district, and subdistrict and then we sorted other imputing variables on activity work, main occupation sector, highest education level, sex, age, marital status, number of household member and dependents. We decided to modify our sequence by putting activity
work first given that we sought to impute labour wages. We believe that the activity work variable would be the first determinant of whether someone earned labour income or not, despite its lower correlation coefficient compared to the main occupation sector variable. The main occupation sector variable was the industry sector where the respondent engaged in their main job. For our last imputing variable, we added dependent from our modified databases.

**Generating the combination of imputing variables**

The next step was to generate the combination of imputing variables. We did this by selecting all imputing variables as linking variables between donor (Sakernas) and master file (Susenas) then sequentially dropping each imputing variable starting from the dependent variable and continuing until the province variable. From this exercise we ended up with hundreds of combinations of imputing variables. We used each of these combinations as a group to form identical cells between the two files to enable a one by one imputation process.

**One by one imputation from donor to master file**

We then started our first imputation process by using all imputing variables. In STATA we merged one to one based on each of the combinations of the imputing variables. We then localised our fully matched observations before we reweighted our Susenas-imputed file.
Figure 4. Step 4 to 5 of the matching process.

Sort variable (ascending):
- Province
- Municipality
- District
- Subdistrict
- Work Activity
- Main Occupation Sectors
- Highest Education Level
- Sex
- Age
- Marital Stat
- No of household Member
- Dependent

Sort variable (descending):
- Income/Salary
- Weight

Main Database
Fully Imputed Susenas 2008

SAK Reserved 2008

Compare weight SAK vs weight SSN

weight SAK $\leq$ SSN
weight SAK > SSN
Subtracted weight SAK - weight SSN
remaining weight SAK

Sort variable (ascending):
- Province
- Municipality
- District
- Subdistrict
- Work Activity
- Main Occupation Sectors
- Highest Education Level
- Sex
- Age
- Marital Stat
- No of household Member
- Dependent

Sort variable (descending):
- Income/Salary
- Weight
Reweighting process

We then reweighted our localised fully matched file. This process was performed by allocating the weight from the donor file (Sakernas) to the corresponding matched master file (Susenas). If the matched observations contained the same weight, we kept this as our fully matched file. Next, if the weight from the master file was higher than from the donor file we kept this as our fully matched file. If the matched observations showed that the Sakernas weight exceeded the Susenas weight, we subtracted our Sakernas donor weight from the Susenas weight. We kept this as our reserved Sakernas donor and we re-used this reserved donor in a subsequent process. Next we allocated the remaining weight to the next matching cell.

Reprocess imputation process

In this step we repeated used our remaining unmatched donor file with the donor reserved file added to impute income variables from the Sakernas donor file to the Susenas master file. An iterative process will finish all of the representative Sakernas donor file by finding observations with the closest income in the representative cells to complete wage information in the primary base file.
Hot-Deck Method

Instead of only using our modified cold-deck to handle the unique case of missing labour income data in Susenas 2008, one of the other common methods to deal with the missing data is the hot-deck imputation method. Hot-deck is a nonparametric imputation method. It does not depend on estimating regression parameters. Hot-deck’s name was derived from the way the survey was initially done manually using survey cards. Originally, the cards were sorted in a way that it was possible to find similar records on the other parts of questionnaire responses. Once the records most similar to the missing records were identified, they could potentially serve as a donor for the missing response. There will be some possibility of an exact match or multiple matched records to fill in the missing values. Using a hot-deck method, the missing values are then filled randomly from an identified series of good matches from the donor set within the same survey data.

With our exercise, firstly we needed to assume that Susenas and Sakernas was one integral survey. So we merged observations with missing labour-income from Susenas before we applied hot-deck imputation. As we did with our modified cold-deck, we then specified the matching variables. In hot-deck, these matching variables were assumed to be correlated with the missing values. Hot-deck next assumed that within the same groups of the imputation cells the non-respondent will follow the same distribution as the respondent (Bechtel, Gonzales, Nelson & Gibson, 2011). Then cases were randomly picked within those groups of the imputation cells as below:
Hot-deck imputation finds and examines the observations with complete records; in our matrix they are obs. 1 and 2. It examines the group of variables between obs. 1 and 2 to determine which observations are most similar for the observation with missing variables (obs. 3). Once found, the substitute value from the most similar observation will be imputed to the missing data point. With this example, obs. 2 is more similar to obs. 3 than obs. 1. Hence our new data matrix after hot-deck will be:

<table>
<thead>
<tr>
<th>Obs</th>
<th>Var 1</th>
<th>Var 2</th>
<th>Var 3</th>
<th>Var 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

After doing this imputation, we transferred our Susenas to their original database and analysed the data using the complete database. For the hot deck imputation we used the STATA package available Mander & Clayton (1999). The inbuilt ‘do’ file in the package processed the Hot-Deck matching procedure to our merged Sakernas with Susenas database.
2.4 Result

Table 11. Summary Statistics: Sakernas raw data compared to Susenas imputed using Modified Cold-Deck and Hot-deck Method

<table>
<thead>
<tr>
<th>Net Income from Main Occupation/Job</th>
<th>Sample of Income Earners</th>
<th>Weighted Population of the Income Earners</th>
<th>Mean (Weighted)</th>
<th>Standard Deviation (Weighted)</th>
<th>Minimum Value</th>
<th>Maximum Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sakernas 2008 (Raw Data)</td>
<td>274,390</td>
<td>60,306,505</td>
<td>947,000</td>
<td>1,269,176</td>
<td>10,000</td>
<td>96,600,000</td>
</tr>
<tr>
<td>Susenas 2008 Imputed Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Modified Cold-Deck Method)</td>
<td>202,422</td>
<td>50,800,000</td>
<td>1,167,272</td>
<td>1,358,512</td>
<td>11,639</td>
<td>75,500,000</td>
</tr>
<tr>
<td>Susenas 2008 Imputed Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Hot-Deck Method)</td>
<td>201,967</td>
<td>47,342,173</td>
<td>1,040,232</td>
<td>1,269,176</td>
<td>10,000</td>
<td>96,600,000</td>
</tr>
</tbody>
</table>

From the results of the general descriptive statistics in table 11, we can see that weighted mean, and minimum and maximum income values from the imputed files using hot-deck matching are closer to the equivalents from the original Sakernas 2008 donor file compared to the results from using our modified cold-deck method.

However, the other indicator, the weighted population show that our modified cold-deck method is preferable. Therefore we further analysed both sets of results using more detailed disaggregations. We conducted a test of the matching process by using geographical areas, income deciles, main occupation sectors and level of education. From there we decided which one is more suitable for our income base file.
An overall comparison to Sakernas donor 2008 (in table 12) shows that our modified cold-deck method preserved a more similar distribution of the imputed income per decile than the modified hot-deck method. Cold-deck results in a lower difference in both mean weighted income and total weighted income for each of the income deciles. None of the results of our cold-deck mean weighted income are significantly different from the decile distribution of the donor. This result could also be further analysed by comparing per decile total weighted income from the Sakernas data. Almost every decile has a similar distribution with the average difference of -20% from the donor files. Deciles one, two, three, seven and eight recorded slightly higher means but with less than one percent differences. Deciles four, five and six recorded more than one percent higher differences compared to the means of the Sakernas decile. Only deciles nine and ten recorded smaller means than the donor (decile nine with one percent smaller mean and decile ten with four percent smaller mean). With a smaller difference in the total income earner per decile and a smaller difference in the total weighted income, the cold-deck method leads to more stable and more similar mean weighted income.
income. Hence our modified cold-deck method performed better than the modified hot-deck counterpart.

Results from an analysis of our modified cold-deck are generally better than for the modified hot-deck imputation. We noted significantly lower coverage of the weighted imputed income from the hot-deck. This also relates to the Table 11 results, where the sample and total weighted number of income earners that can be imputed using the hot-deck method are far lower than for the modified cold-deck. The same pattern of the mean imputed income can be seen from the result of hot-deck imputation. While the number of imputed income earners is lower than for the modified cold-deck, we can see that, compared to the Sakernas donor, hot-deck consistently provides a lower mean weighted income in every decile except decile eight. Thus the overall result of the mean weighted income and total weighted income differences favour the modified cold-deck method.
Rubin (1996) emphasized the limitation of imputation approaches in their inability to recreate individual values. Therefore, the assessment of the imputation methods could be achieved in terms of the preservation of important distributional aspects and the multivariate
relationships. Table 12 shows how our modified cold-deck method performs better in preserving the mean income distribution within the decile of the new base file.

In table 13 we attempt to further compare the regional level distribution of matching results. We noted that with the consistent national bottom line, our modified cold-deck method performs better than the modified hot-deck. In addition, we also need to see how the provincial means of our imputed data differ from the donor. By using our modified cold-deck method, 7 out of 33 provinces showed a significantly different income with more than 10 percent difference of the mean weighted income compared to Sakernas donor. This result is far better than the result from the hot-deck method which could preserve mean weighted income from 12 out of 33 provinces with less than 10 percent difference to its donor.

From the analysis of the total weighted income, we conclude that the bigger difference is most likely associated with the smaller number of total imputed income earners in the Susenas database. We were unable to do any modifications to improve this. Our matching procedures only aimed to fill the missing labour-income for income earners from the Susenas observations. However, the results of the mean-weighted income show us a relatively small distributional difference. Hence, looking at table 13, we can confirm that our cold-deck method preserved a more similar distribution to that of the donor than the hot-deck method.
One limitation of the hot-deck method is the missing imputed people with a postgraduate education. In addition, Table 14 shows that the mean weighted income from the hot-deck approach is distributed differently compared to the donor’s mean weighted income. Mean weighted income should have increased in line with the increase in the education level of the respondents as shown by the Sakernas raw data 2008 and the result from our modified cold-deck method. However, this is not the case with the hot-deck method.

Despite the similar distribution of the mean of weighted income shown by the cold-deck method in Table 14, the preservation of the distribution of the total weighted income and mean weighted income based on the level of education is not as good as our previous disaggregation result by decile and by region.
Additional information to note on the analysis from Table 15 is that our Susenas base file has slightly different characteristics compared to the Sakernas donor. After the harmonisation of matching variables for all the observations of Susenas and Sakernas, we compared the distribution of main occupation sectors in both databases as shown previously in Table 8.

Table 8 confirms that the estimated population from the two different sources includes all the individuals who work in both the unclassified and classified sectors. Further, from that harmonisation we omitted workers without income from our Sakernas donor before the matching process. In other words, we only used income earners from the donor to be imputed to Susenas. In Susenas we also omitted the non-income earners identified in the data and imputed the income earners flagged as having missing labour income variables.
We also found that there were no income earners donated from Sakernas working in unclassified sectors; however, there were income earners who work in unclassified sectors in our Susenas data. Therefore table 15 confirms the end result of our matching test where there are income earners with imputed labour income in unclassified sectors from both the cold-deck and Hot-Deck results. In short, we note that for the detailed results our cold-deck method still performed better than the hot-deck method. We could observe that with the cold-deck, the distribution of mean weighted income from six sectors is similar to the donor. However, in line with Table 8 which noted a big difference in the distribution of the total weighted observations between utilities (electricity, gas and water) and financial, real estate and rental services sectors, the results in table 15 are similar.

2.4.1 The Remaining Gap Problems

Although the data matching imputation method could provide the Susenas data with income from employment salary/wages, we need to acknowledge that both the Susenas and Sakernas surveys still failed to cover the higher income households. As previously explained in Chapter I and endorsed by similar experience from more advanced countries, we need to maximise the advantage of available survey and administrative dataset in developing the microsimulation base file (Paas, 1986; Atkinson, 1997; Decoster and Van Camp, 1998; Wagenhals, 2004; Lloyd, 2007; Eurostat, 2013). This study tries to overcome the issue of Susenas and Sakernas not covering the top income earners by using tax return data to augment the upper income groups. It is important to note that the ability of the tax return dataset to provide information on the upper income households depends on the proportion of upper income households that lodge their tax form.
The process should first identify the part of the tax return that is covered by the Susenas-Sakernas income data. This means we need to link variables from the tax return data to the Susenas-Sakernas income data. Since the tax payer characteristics variable in this dataset is not as rich as the individual characteristics information in the survey data, the possible linking categories are the geographical category, the income level category and the dependent category.

Initially, we matched the geographical information in the tax return based on the location of the receiving tax office to the Susenas provincial code. After that, we grouped cells of 17 pre-determined income brackets, ranging from 25 million rupiahs to 3 billion rupiahs of gross annual income. This income bracket is based on the income range that has different tax rates applied as well as the lower and upper income range coverage by the PIT Return and Susenas. Finally, by comparing those two group cells, we carefully identified income earners not covered by our Susenas imputed dataset but that could be imputed by our PIT return dataset. This process added 2.2 million observations to the estimated population in our dataset. Although this number is less than 5 per cent of the overall income earners (approximately 57 million), it added more than 50 per cent of the income and the database was now ready for simulation. In figure 5, we plot the annual income salary distribution with the weighted population from 5 sources of data; Sakernas 2008, Susenas by using hot-deck imputation, Susenas by using cold-deck imputation, Individual Tax Return 2008 and lastly the Susenas with cold-deck imputation augmented with tax return top income.
It can be seen from the plot in figure 5 that cold-deck approach covers more income earners than the hot-deck estimation, but hot-deck could reach higher income earners better than cold-deck. At the bottom, with the smallest coverage of the income earners, is the tax return data that consists of around 4 million taxpayer returns. Tax return encompasses Sakernas and the two imputed base files for the coverage of the higher income earners. Thus we sought to increase the coverage of higher income earners on the base file by adding the tax return information to our Susenas modified cold-deck dataset.
The kernel density estimate in figure 6 shows the mode of the labour salary between the green and red arrow lines. The first vertical arrow-line is the lowest minimum regional wage for 2008 (which is 6 million rupiahs annually for East Java). The second vertical arrow-line shows the highest minimum regional wage of 13.27 million rupiahs annually which is recorded for both Papua and West Papua. The mode from the tax return plot is somewhat higher at around 25 million rupiahs annual income. We might expect that most of the employees and labourers from both Sakernas and Susenas imputed by Sakernas income base file are under the threshold of taxable income. The allowance for a single person’s non-taxable income is 13.2 million rupiahs based on tax law 17/2000. Hence, it is important to increase the coverage of higher income earners from the tax return data.
By imputing tax return data to our Susenas modified cold-deck dataset, we increased the coverage of the high income earners and we could match the figure of the actual personal income tax (PIT) national data. This calculation used our PIT estimator which will be explained in detail in the next chapter.

2.5 Conclusion

We have met the challenge to improve the Indonesian income data base in order to establish a better basis for microsimulation. Later we will use this better database to build a personal income tax microsimulation model. A much better income database for modelling Indonesian tax microsimulation has been produced. This model serves as a strong basis for
the study of the impact of tax reforms; more specifically, to assess whether the reforms will be capable of generating a source of sustainable public financing.

It is usually not common for a researcher to engage with the imputation of the primary original data. It is usually the survey institution or the statistics authority attempting to complete missing information in the surveys. The fact that the 2008 Susenas dataset did not have published data for labour income made us try to solve this problem. We explored and assessed the possibility of using Sakernas 2008 as our labour income donor. By modifying two particular methods of imputation to handle this problem of missing data, we have shown how to create a more complete database for Indonesian microsimulation modelling.

Due to its long history and continuous improvement and use in previous studies, we decided our experiment would deal with the missing data of Susenas by using the counterpart Sakernas information. This method of using another base file as a donor is commonly known as a cold-deck method. We also employed a modified hot-deck method to compare to the cold deck method. This method first assumed that Susenas and Sakernas were one group of data and then imputed existing variables to the missing ones.

The methodology and implementation guidelines for statistical matching are from the European Commission on Statistics (Eurostat) 2013. We completed the first step in harmonisation and reconciliation of both our donor and recipient base files in order to enhance the statistical matching process. We have also analysed our results using disaggregated data based on the harmonised main matching variables. Our general
assessment of the imputation result led us to identify the cold-deck modified method as the better approach to handle the problem of missing income data. The cold-deck imputation completed income information for labour income earners in the Susenas dataset. Most of the significant differences from the imputation were due to the difference in number of the sample observations and the weighted populations or, in other words, the different number of income earners, between the donor (Sakernas) and the recipient (Susenas). The bigger the difference between the two components, the bigger the absolute difference between total imputed labour income (in Susenas) and total labour income (from Sakernas donor). However, the distribution of mean imputed labour income (in Susenas) is generally similar to the distribution of mean real labour income (from Sakernas donor). Hence, we were assured that the distribution from the donor is preserved by our modified cold-deck method.

In the final section of this chapter we acknowledge that the coverage of our imputation base file is less than the donor file. This result applied to both the total income earners and to the sample coverage of higher income earners in the base file. Subsequently, to solve the problem of under coverage of higher income earners, we compared the plots of Sakernas, Susenas imputed, and the personal income tax return database. This confirmed the need find a way to add the higher income earners tax returns, thus addressing the previous concerns about the lack of coverage of high income earners which has been considered an inherent weakness of the available Indonesian income micro data. To solve the problem, the high income earners from tax return data were imputed to the Susenas’ imputation base file. Our validation results confirm a significant increase in the coverage of the higher income earners. However, we have another highlighted and pending problem of the lower number of income earner observations from our Susenas imputation. This lower number of income earners will lead to a lower number of eligible taxpayers who can be detected from the imputation base file. In
the next chapter we will discuss our solution to improve this database by reweighting the income earners to match the number of income earners reported in the BPS publication (2009).
CHAPTER III. DEVELOPING INDONESIAN LABOUR INCOME TAX MICROSIMULATION

3.1 Introduction

For more than three decades microsimulation of the income tax policy has been continuously developed and enhanced by the rapid development of computing technology. Microsimulation has proved to be a powerful tool to estimate the revenue and distribution impact of tax policy, especially in developed countries (Gupta & Kapur, 2000). Further, many developed countries have made microsimulation models their standard evidence based-tools for fiscal policy design and evaluation (Bargain, 2007). The use of microsimulation began in 1957 when Guy Orcutt pioneered it as a tool for economic analysis.

Microsimulation compares the observed state and behaviour of a sample of economic agents (individuals, households or firms) before and after a change happens to their environment (Bourguignon & Spadaro, 2006).

This chapter will focus on the development of microsimulation tools for Indonesian labour income tax. A sequential step by step explanation of the development of a tax microsimulation model will be elaborated in detail. Microsimulation performs ex-ante quantitative analysis to evaluate the possible impact of fiscal policy change (which in our case is the labour income tax policy change) prior to implementation. The strength of this model lies in its ability to include whole characteristics of the economic agents within the heterogeneous sample of the individual micro-data. With this ability, the model is capable of
giving a detailed overview of the tax reforms’ distributional impact among the individual economic agents in the observations. Finally, after the individual impact is captured it is easier to calculate the financial costs and benefits of the reforms as an aggregate (Mitton, Sutherland & Weeks, 2000, Bourguignon & Spadaro, 2006).

3.1.1 Tax Microsimulation in Indonesia

Microsimulation models in developing and transitional economies are still relatively rare and of limited use. Most of the existing models for developing and transitional countries are still in their very early stages of development. As mentioned in section of review of literature in chapter I, for Indonesia, there are at least two examples of unpublished studies using tax microsimulation in particular. Yuwono (2008), in her PhD dissertation, used micro-level data from salary-income withholding tax returns from the Indonesian Directorate General of Taxes (DGT). Using administrative data, her personal income tax (PIT) microsimulation examined the distribution of income tax burden and revenue impact from several scenarios of personal income tax reform. Her results suggested a trade-off between revenue impact and the distribution of the tax burden.

The second unpublished study was undertaken by Marks in 2003. Different from Yuwono (2008) who used administrative data, Marks used the National Socio-Economic Survey (Susenas) 2002 to calculate income and income tax obligations of Indonesian households. He conducted the simulation based on Income Tax Law no 17(2000). His approach was the first detailed micro data simulation for Indonesian personal income tax based on the survey data.
However, he did not take into account that the unit of Indonesian income tax, as stipulated by law, is the individual tax payer unit. Instead, he used household as his unit of analysis. He highlighted significant gaps between potential revenue and government revenue collection from personal income tax.

From these two Indonesian PIT models, we commenced our attempt to improve the outputs by taking some key learnings into consideration. We developed a more detailed microsimulation model for Indonesian personal taxpayers by using an approach that combined the power of the available income micro data in Indonesia as described in Chapter II.

3.1.2 Utilisation

The improvement of the national statistical office and other survey institutions, together with the growing public awareness and increasing demand for more reliable data and accountable policy processes, are among the factors that have provided an adequate supply of detailed information to support public debate on tax policy initiatives. In addition, major improvements in the socio-economic research methodology and the computing technology have made possible a much greater depth and breadth of analysis of the policy process. At almost every step of public policy analysis in developing countries, be it for the initial stage of the legislative initiatives, the development of policy, or in the analysis of the successes or failures of the enacted programs, microsimulation is among the most frequently used tools to facilitate the analysis. Its unique ability to better replicate the heterogeneity of the population
and to provide more transparent analysis of the inter-relationships between population subsets has meant that this modelling technique is in high demand (Buddelmeyer, Creedy & Kalb, 2007).

Thus tax policy design and analysis can be improved by the use of microsimulation techniques. Microsimulation enables detailed ex ante studies of proposed or actual tax reforms. There are two types of microsimulation model which have been used by the governments of developed countries; static and dynamic models. The static microsimulation model is most often used to examine the potential impacts of detailed changes to tax and transfer policies. The dynamic microsimulation model is often used to analyse long-term and behavioural impacts. The static microsimulation model assumes that there is no behavioural change from individual micro-units as a response to the policy changes or other events. The dynamic model, on the other hand, tries to estimate some of the behaviour of individual micro-units over time as a response to the policies and/or to the demographic and economic events (Lewis & Michel, 1990). Dynamic microsimulation models age the dataset over time and simulate the change of each attribute in each individual micro-unit over time (Li & O’Donoghue, 2013). Harding (1996) explains that the change in the attributes is constructed to follow the probabilities of major life events, such as marriage, divorce, birth, fertility, death, education, migration and labour force participation, for each of the individuals in the dataset. To sum up, dynamic microsimulation recalculates an individual’s attributes for each episode in the simulation period.

Static microsimulation models do not age the dataset over time and do not model life events. They assume that the probability and proportion of people married, divorced, migration,
death, born, in the labour force and in education on the base dataset do not change. Static microsimulation models are used to simulate rule-based systems by replicating the set of rules and applying them to certain attributes of the individual micro-units. In the specific case of tax policy studies, static microsimulation recalculates the eligibility of micro-units to pay taxes based on the set of stipulated tax rules/law.

The capacity of microsimulation to simulate the conditions of individual micro-units before and after policy implementation has made it a powerful tool for examining the distributional impact of many underlying public policies. Since its initiation, these models have been intensively developed for more than three decades. There are many studies on the application of the microsimulation model to the analysis of tax-benefit reform. Provided the needed variables and consistent panel information are available on the survey dataset, it is also possible to undertake behavioural modelling when tax reform is specifically aimed at influencing tax payer behaviour.

Atkinson, King & Sutherland (1983), Citro & Hanushek (1991), Eason (1996), Naylor (2000), Wagenhals (2004), Granell-Perez, Fuenmayor-Fernandez & Higon-Tamarit (2006), Lloyd (2007), among others, exhaustively described, summarized and evaluated recently used tax-benefit microsimulation models in several developed countries including the United Kingdom, the United States, Canada, Germany, Spain and Australia. The various studies and evaluations covered the implementation of tax-benefit models in advanced welfare-states on three continents; Europe, North America and Australia.
The development and improvement of microsimulation as a tax policy analysis-tool has involved academia, government and the public and has enhanced interaction between them. Government counterparts are usually represented by the Inland Revenue Authority or Ministry of Finance. In some countries, the Ministry of Social Welfare/Social Security Agency and the Treasury are included. Public involvement is in the legislative debate with the congress or parliament and through the dissemination of the policy initiatives, processes and results by the media. Therefore some studies argue that the need to make well-documented and transparent models is the main agenda for the improvement of these models.

Amongst the latest and most important milestones has been the development of a cross country microsimulation model. This was marked by the start of EUROMOD, a European Union wide microsimulation model which integrated fifteen EU countries and aimed to overcome different datasets and modelling assumptions to enable analysis at the national level and multi-country EU level (O’Donoghue, Sutherland & Utili, 1999).

In discussing the choice of the model to be built, Mitton, Sutherland & Weeks (2000) emphasized the need for the policy question to be addressed. In the previous chapter, we elaborated our attempt to maximise the usage of current available data to construct an Indonesian income base file which can better capture the income distribution in Indonesia. By using the previously developed base file, this chapter focuses on building a tax microsimulator, a static personal income tax model and testing the model based on its performance in addressing the main research questions.
3.2 Personal Income Tax (PIT) in Indonesia

To start with, any microsimulation model is going to depend on the definition of the simulated variables. This section will discuss the crucial definition of income for the Indonesian PIT microsimulation model.

3.2.1 Definition and Coverage

Under the current Income Tax Law 36 (2008), income is defined as any additional economic ability (income) received or earned by the taxpayer, either from Indonesia (domestic income) or outside of Indonesia (international income), that can be used for consumption or as additional wealth to the taxpayers in any name and in any form. The definition of individual or personal income includes labour income or income from employment in the forms of salary, wages, honoraria, bonuses, commissions, gratuities, pensions, and other form of income from personal business profits.

Since the inception of the tax reform in 1983, Indonesian income tax law adopted most of the taxation concepts from more advanced countries (Gillis, 1985; Heij, 2001). There was involvement by experts from the World Bank, the International Monetary Fund (IMF), the German Development Agency, and technical assistance from expert teams from the Harvard Institute for International Development. Therefore the Indonesian income concept and definition in the personal income tax law is a modification of the Haig-Simons definition of income. Some economists view Haig-Simons as an optimal definition of income for tax
purposes. Haig’s (1921, p. 27) definition of income is “the money-value of the net accretion to one’s economic position between two points of time”. A more specific definition of personal income by Simons (1938, p. 50) as “the algebraic sum of (1) the market value of rights exercised in consumption and (2) the change in the value of the store of property rights between the beginning and the end of the period in question”. Rosen (2002) shortened this definition to a net increase in individual’s consumption ability of the goods and services any additional wealth from any sources. Some argue this is the optimal definition, but some others are concerned about it being overly broad since all income from any sources is included. Further, Rosen (2002) argued some of the problems from the Haig-Simons definition includes difficulties in measuring capital gains and losses (realized and unrealized), in-kind services, and imputed income from durable goods.

The detail of articles in the Law no 36(2008), including detailed coverage of the income tax objects per article is shown in table 17.

Table 17. Article and Object of Personal Income Tax under Law no 36(2008)

<table>
<thead>
<tr>
<th>Article</th>
<th>Income Tax Object</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article 21</td>
<td>Withholding income</td>
<td>Includes: wages, salary, commissions, bonuses, allowance, pensions, gratuities, Honoraria, lottery prizes, awards, Insurance premiums</td>
</tr>
<tr>
<td>Article 22</td>
<td>Corporate income tax for specific industries</td>
<td>Profits from business</td>
</tr>
<tr>
<td>Article 22 import</td>
<td>Income tax from imported goods and services</td>
<td>Taxed on cost, insurance and freight</td>
</tr>
<tr>
<td>Article 23</td>
<td>Interest, dividend, rents from property, royalties</td>
<td>Includes: annuities received or accrued</td>
</tr>
<tr>
<td>Article 25/29 (individual and corporate)</td>
<td>Real income of individual/corporate at the end of tax year based on the amount of tax owed last year</td>
<td>Includes: Profits from business, refunds of tax payments which already deducted as cost</td>
</tr>
<tr>
<td>Article 26</td>
<td>Income tax for expatriate either individual or corporate from income sources in Indonesia such as interest, dividend, royalty, rent, etc.</td>
<td>Profits from business</td>
</tr>
<tr>
<td>Article 4 (1) Final Tax</td>
<td>Any income sources that are already taxed at the final stage, no need to report this under the individual income tax form.</td>
<td>Includes: gains from sale or transfer of property, gains from the difference in exchange rates, gains from cancellation of debts</td>
</tr>
</tbody>
</table>

Table 17 shows the modifications introduced by the Indonesian government to the Haig-Simons concept of income for personal income tax. Indonesian tax considers the changes in
net wealth including wealth from realised gifts, bequest, and other gratuities transfers. But, as applies in most countries, Indonesian tax does not recognise some changes in net wealth which are not yet realised, such as the form of income from imputed rental value of owner-occupied homes.

### 3.2.2 History of IT-Based Tax Records in Indonesia

Before the 1983 tax reform, the German Development Agency initiated technical assistance to reorganise and computerise part of the Indonesian tax system (Heij, 2001). By late 1989, the personal income tax data had just started to be available in electronic format (Leigh & van der Eng, 2009).

DGT began to develop its first electronic-based information systems in late 1989. It marked the start with an application to monitor and evaluate tax payment called NPCS (Newly Payment Control System). In 1994 DGT made improvements to replace NPCS with the development of a taxation information system (SIP-Sistem Informasi Perpajakan). SIP is a local automation of NPCS which accommodates all taxes except land and building tax and transfer duties. Different from the NPCS which only dealt with tax payment, the SIP monitored both tax payment and tax return submission. This function was adopted because the Indonesian tax law required the taxpayer to pay tax before they can lodge a tax return. In addition, SIP also served as a decision support system whereby it provided data and information from the submitted and recorded tax return. The analysis was limited to the submitted and recorded data provided by SIP, with the SIP reporting and recording processes
both still manual and done in separate sequences. Another type of tax information system management was introduced for land and building taxes. It especially applied to the Land and Building Tax Office.

Over the years, some other improvements were made by the DGT. The addition of wide area networks and satellite-wide systems to enable inter-office and regional-areas administrative data transfer were introduced in 1998. Further, at the same time as converting 6 digit dates to 8 digit date format for Y2K (year 2000), DGT started to develop an application for Value Added Tax (VAT) in and out. This was done to foster the efficiency of inter-office VAT reconciliation. It also aimed to lower the audit burden, as by law VAT tax auditors are required to cover full transactions of VAT audits. This year also marked the initial development of future information systems and a technology blueprint for DGT.

The first regulation for the submission of the PIT electronic return was initiated in 1999. The use of electronic media as a replacement of form 1721-A1 was stipulated by the DG of Taxes Decree No. KEP-13/PJ./1999 (18 January 1999). This is a secondary form, which is an attachment specifying a detailed breakdown of income, allowance and expenses for inclusion in the main form of Income Tax Return article 21 (Form 1721). Further operational details for this policy were circulated to tax offices with DG of Taxes circular letter No. SE-03/PJ.43/1999. There is a precondition for tax payers to be able to get a permit to electronically submit the attachment to the master file of the withholding tax return. Taxpayers are required to apply a computerised payroll system previous to the request for electronic returns’ submission. Requests should be directed to DGT and must be accompanied by computer records in the form of a disk. All payroll data must be translated
into a text file format from their original accounting and tax systems format. DGT’s centre for taxation data and information will check the compatibility of the taxpayers’ data with the DGT applications program. If it is compatible, a letter of approval will be issued within 30 days from the submission of the request. However, based on this ruling, the electronic submission was only applicable to 1721-A1 (which is a breakdown detail of salary for private sectors employees). The other appendix/attachments to the master file of form-1721 must still be submitted using its original form with a signature of the taxpayer or authorised person on behalf of the taxpayer on the form.

DGT finally launched the full electronic tax return formally at the beginning of 2008, following Ministry of Finance Regulation No. 181/PMK.03/2007. It was not until then that DGT started to provide an e-Filing application. This application enabled electronic tax return submission online and in real time. At first it was provided through Application Service Provider (ASP) companies. In 2009, DGT improved the service by providing a section for e-Filing on the official website of the DGT. Besides the introduction of electronic filing and electronic returns, DGT has made several other improvements to the way it collects tax returns. During the last month of the fiscal year, tax payers can choose to submit their manual/hardcopy return to several locations. They can submit to a one-stop shop section at the tax service office, to a tax corner in a trading centre, like a mall or traditional market, in every city in Indonesia. In addition to that, some regional offices launched tax-mobile cars, a mobile service that goes around particular cities and collects tax returns during the tax return submission period. Further to the mobile-services, several points of tax return drop-off boxes in some alternative public places other than the tax offices location were also provided during the tax return submission-period. This advancement of services was expected to increase the compliance rate of tax return-submissions. The expected end result will also support the
completeness of the electronic tax return databases. This better database is expected to support better policy analysis and better decision making processes.

Apart from providing these electronic-based services to assist taxpayers in complying with tax filing requirements, DGT still maintained rigid criterion of a complete tax return to ensure the quality of electronic returns submitted by taxpayers. The tax return must be completely correct before the tax office could accept the tax return forms. With the application of the 2007 electronic filing and electronic return regulation, the tax return is categorised as complete when all elements of both the master file and the attachment form are completed. The master or main tax return must be signed by the taxpayer or the authorised person. An electronic signature is permitted upon request and proper authorisation from the DGT. The electronic signature must be completed with its specific attachments and required information and/or documents. This applies for both manual and electronic returns, except that for the electronic return all related documents which cannot be reported electronically must be sent or submitted directly to the related tax office following the submission of the electronic return through e-filing. The acceptance of the tax return leads to the tax return recording process. In the imputation process explained in the previous chapter, the donor of higher income earners were the 2008 recorded personal income tax returns from the DGT system. Compared to the previous years’ data, the quality of the 2008 recorded returns was a lot better due to the increasing compliance rate of tax return submission in 2008 (DGT, 2009). However some missing variables and information, and also some inconsistencies in the filled columns of the return were still found. Thus a consistency test and a reasonableness test for each record was performed and outliers were not included in our base file.
3.3 Methodology

3.3.1 Scope and Ruling

The scope of our built PIT model is slightly different to previous studies on Indonesia by Marks (2003) and Yuwono (2008). Our personal income tax microsimulation focuses solely on the labour-income tax from the provision of Income Tax Law no 36(2008). Using Susenas 2002, Marks (2003) built a microsimulation model for all types of the personal income tax component under Income Tax Law no 17(2000), namely article 21 for income from employee/salary earners, article 25 for income from personal business, article 4(1) final tax for income from rental, some financial transactions and capital income, and lastly article 23 for income from interest, dividend, royalty and prizes. He used STATA to prepare the underlying database for the model. His database is a combination of the core module and the income-expenditure module from Susenas 2002. He then used visual basic language to code the application of PIT from the interpretation of Law no 17(2000) to all household units in the Susenas 2002. While Marks simulated all types of PIT from Law no 17(2000) to household respondent in the Susenas 2002, Yuwono (2008) used administrative data from the previous years’ PIT returns to provide ex-ante analysis of the major 2008 Indonesian PIT reform. She used STATA to code the scenario of the 2008 PIT reform to apply it to sample of individual tax returns from DGT personal tax administrative data.

Bourguignon and Spadaro (2005) elaborated three key factors of microsimulation models for redistribution analysis, namely; 1) a micro-dataset of information on the economic agents (individuals and/or households); 2) a set of public policy rules; and 3) a theoretical model of
the economic agents’ behavioural response. Our model enhances the scope of the first key factor; a micro-dataset. To improve the study and analysis of previously built Indonesian models, we combined both the power of national survey data and tax administrative data. Details on the methodology were elaborated in Chapter II of the thesis. This chapter focuses on the second key factor of building a labour income tax microsimulation model by coding the rules of labour income tax from income tax Law no 36(2008). We used STATA as our main programming language for both managing the database and for coding the set of rules.

3.3.2 Variables and Determinants

For the personal income tax microsimulation we need to analyse all of the available variables and determinants in the base data. Variables were derived from all available attributes from each respondent in the individual micro-units. In addition to that, determinants are the important and relevant variables to be included in the income tax microsimulator. Determinants will be employed to recalculate the personal-labour income tax liability based on both the applicable tax code and other tax code possible scenarios. This section of the thesis identifies the variables and determinants from Sakernas 2008.

We used administrative tax returns as the main benchmark source of variables to build a tax simulator for the Indonesian labour-income tax microsimulation model. Based on the details of the administrative tax returns, there are two groups of variables from the personal income tax return; individual information and income details for tax calculation. Below we precede our explanation of the model by tabulating these two groups of variables as the main
benchmark to identify similar variables provided by both our Sakernas donor survey data and Susenas cold deck imputed survey data to be used in the tax simulator. We also discuss our approach to including possible variables from both surveys and explain our adjustment to the variables.

As argued by Atkinson (1997) and Paas (1996), generally it is difficult to find specific household surveys that match with the variables needed for tax microsimulation. We are required to adjust the available variables in the survey to prepare them for our tax simulator. Table 18 above shows the result of the comparison between variables and determinants of the tax simulator in the tax return and in the Sakernas survey. It can be seen from information on the individual characteristics that Sakernas has both determinants for the exemptions in the tax simulator. Thus, marital status and number of household members were used to allocate the dependents to the tax payers from the Sakernas sample of observations. Note that the dependents in income tax law are different from the number of household members available in the survey data. There are specific requirements for the dependent in the income tax law. We elaborate this in the next section on the adjustments needed to find the dependents from our survey base file.
Table 18. Determinants from PIT Return and SAKERNAS variables’ mapping

<table>
<thead>
<tr>
<th>TAX RETURN</th>
<th>SAKERNAS (*)</th>
<th>REMARKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual information:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>name</td>
<td>n/a</td>
<td>unidentified/anonymous</td>
</tr>
<tr>
<td>tax identification number</td>
<td>n/a</td>
<td>unidentified/anonymous</td>
</tr>
<tr>
<td>address</td>
<td>n/a</td>
<td>unidentified/anonymous</td>
</tr>
<tr>
<td>marital status</td>
<td>√</td>
<td>key variable for calculating non taxable income</td>
</tr>
<tr>
<td>sex</td>
<td>√</td>
<td>available in both databases</td>
</tr>
<tr>
<td>(foreign) employment status</td>
<td>n/a</td>
<td>Sakernas information only covered domestic workers, no foreign workers identified from the data</td>
</tr>
<tr>
<td>number of dependents</td>
<td>√</td>
<td>key variable for calculating non taxable income</td>
</tr>
<tr>
<td>job’s position</td>
<td>√</td>
<td>available in both databases</td>
</tr>
</tbody>
</table>

**Income detail for income tax calculation:**

**Gross income:**

- salary/pension | √ (*) | gross-up salary was calculated in chapter 2. (*) No information about pension payment |

**Deductions:**

- income tax allowance | n/a | not available in SAKERNAS |

**Income in the forms of goods/services** | √ | included in total salary |

**Insurance premium paid by employer** | n/a | not available in SAKERNAS |

**Bonus, gratification, production services and special religious festive allowance** | n/a | not available in SAKERNAS |

**Income tax calculations:**

- o net income; | √ | |
- o previous net income; | n/a | not available in SAKERNAS |
- o previous net income for income tax calculation (annually/annualised); | n/a | not available in SAKERNAS |
- o non taxable income; | √ | calculated using tax estimator |
- o taxable income (annually/annualised); | √ | calculated using tax estimator |
- o income tax from taxable income (annually/annualised); | √ | calculated using tax estimator |
- o income tax liabilities; | √ | |
- o income tax borne by government; | n/a | not available in SAKERNAS |
- o income tax to be withheld; | √ | calculated using tax estimator |
- o income tax withheld/paid previously; | n/a | not available in SAKERNAS |
- o over-payment / under-payment of income tax | n/a | not available in SAKERNAS |

**Note:** SAKERNAS (*) is a donor to our Susenas imputation, therefore the basefile has the same limitation as Sakernas and the same adjustment process to the variable in Sakernas need to be applied.

Further, we assumed that the available information on the labour/employment income from Sakernas is sufficient to be included in our tax simulator. We expect that some details will be
missing from our tax simulator, such as the variables relating to foreign taxpayers, pensioners, and non-permanent employees. There are also some other missing details on the gross income components and allowances. We assume that the grossed-up labour income from the Sakernas is an accumulative number of all components of gross income required for the PIT calculation.

### 3.3.3 Adjustment and Formula for PIT Simulator

This section describes details of the adjustments applied to the available Sakernas survey data to enable it to be used for the tax base data. It starts with the important adjustment of the tax unit then continues with the allocation of dependents from the household in the survey to the income earners (tax payers) in the family. Once dependents are allocated, we proceed to recalculate eligible exemptions. Further to that is the treatment of different income units in the survey. The final step after the adjustments is the application of the tax simulator formula.

**Unit of observations (Tax Unit)**

Before we proceed to our formula for PIT simulator, we need to understand the important features of PIT article 21 calculations. The main thing to be considered when using income survey data is its different unit of observation compared to the administrative data from income tax returns. The income unit in the survey data is totally different from the tax unit in the tax return. Our raw base file Susenas income module contains household-based micro data while the Indonesian personal income tax unit is based on the individual unit tax payers. Therefore we adjusted for this difference by transposing the information from the household
level in Susenas income-module 2008 to the Susenas core-2008 individual unit. We have already done this (Chapter II) to meet a precondition for our first base file before we conducted the two imputations; the imputation of labour income from Sakernas and the imputation of higher income earners from PIT return data. For this imputation, the Sakernas sample observations’ unit and the tax return unit need no adjustment as both are individual-based data.

**Allocation of Dependents for Eligible Exemptions**

There are two forms of most generally known tax relief, tax deductions and tax credits (Messere, De Kam, Heady, 2003). A tax deduction (also called tax allowance in some other countries) is a lump sum deduction from gross annual income. This is the threshold for the non-taxable income, below which there is no PIT payable. This non-taxable income is a function of the tax payers’ status of employment and the tax payers’ dependents. This lump sum deduction is the second important feature that should be adjusted based on the household characteristics of the micro data from the survey. From the members of the household we needed to find possible dependents to be allocated to the individual tax payers in the data. The allocation was based on several rules as stipulated by the law. Note that by law, tax payers are considered as individuals not as families or households. There is only one taxable income threshold for all individual tax payers, regardless of their marital status. Additional exemptions are available for married individuals and for individuals with dependents. The rule is as follows. Additional exemptions for the same dependents can only be claimed by one taxpayer, for example if a father or first income earner in the family has already claimed deductions for his children, then the mother or any second income earner can claim exemption only for his/herself (with no dependents). Hence, if a husband has already claimed
an additional exemption for being married, the wife can only claim herself. She is not entitled
to claim her marital status in the exemption (i.e. a working-wife who earns money and files a
tax return which is not combined with the tax return from her working-husband will not be
entitled to claim status of her marriage in the component for taxable income exemptions).

As mentioned in the previous section, after completing the individual base file, it is crucial to
identify and allocate dependents from the household to the individual income earner who
meets the criteria as the tax payer in the household. Determinants of non-taxable income
entitlement in Sakernas data are: information on relationship to head of household, sex, age,
marital status and number of household members. In the first instance, we identified the
income and non-income earners in the household using the income tax law criteria. We then
identified possible dependents from non-income earners in the household. Those were all
persons under 18 years old or a person over 18 years old but earning no income and not
married yet. These possible dependents were then allocated to each of the income earners in
the household. For this allocation, each of the income earners in the household was entitled to
deduct a maximum of three dependents following the income tax rule. The basis of our
dependent allocation was the hierarchical relationship. We started from the head of
household. If he/she was an income earner then he/she was the first person allocated the
maximum number of dependents. After that we then moved to the other income earners in the
household for the allocation of the remaining possible dependents. On allocation of
dependents we assumed no joint returns and so a working wife or woman in the household
will only bear herself. There is an exception in the law that if the woman income earner is the
head of household then she is entitled to a maximum of three dependents. The treatment of
joint returns and individual returns are both permitted by the law. There is no difference in
tax exemptions for the final calculation of income tax liabilities. Be it joint or individual
returns, the maximum dependent entitlement is only applicable to the head of the household.

**Income unit assumptions**

In line with the previous treatment for allocating dependents, we also assumed there was no
income combination or joint return reporting from the household. This means that we treat
each income earner in the household as a separate taxpayer who files their own tax return.
We omit the possibility of a combined working-wife and husband income. Most importantly,
both of the treatments are permitted by law. We have to limit our simulations to avoid
unnecessary additional assumptions which could complicate our base data.

**Tax simulator Formula**

After the built-in assumptions and the main income dataset were completed, we proceeded to
the second component of the model, which is a parameter dataset for our tax simulator. We
followed a previous Indonesian PIT static microsimulation model developed by Yuwono
(2008) with some modifications.

The tax simulator details all of the Indonesian tax structure variables as stipulated by the tax
law. It starts with the calculation of taxable income from the annual gross income minus non-
taxable income. Different from Yuwono (2008) who used PIT returns, we used a combination
of survey and administrative data. By the use of survey data, our gross income was calculated
using an iterative process while our taxable income is a function of the number of dependents
based on the personal exemptions simulation. The details of the preparation of our base file was explained in the previous chapter. The simulation produces income tax liability from taxable income multiplied by the marginal tax rate.

**Figure 7. Schematic of the PIT structure in Indonesia for the calculation of 2008 tax returns (Based on the Law no 17(2000) and Ministry of Finance Decree 37/PMK.03/2005)**

<table>
<thead>
<tr>
<th>Gross Income(-)</th>
<th>Pension/Occupation Allowance 5% of gross income, or max IDR 1,296,000 annually</th>
<th>( = )</th>
<th>Annual Net Income</th>
<th>( - )</th>
</tr>
</thead>
</table>

**Personal Income Tax Exemptions:**
- Single: IDR 13.20 million
- Married: additional IDR 1.20 million
- Combined income: additional IDR 13.20 million
- Dependents (maximum of 3): additional IDR 1.20 million

( = )

Taxable Income

### 3.4 Application of Tax Microsimulation in the Indonesian Income Tax Policy

This section explains the application of the microsimulator on several base datasets. By applying the tax simulator into several different databases we aim to compare the result from each of the databases. From a comparison of the results, we can see the available and missing components. Thus we can identify the benefits and the limitations of using each database for tax microsimulation.
3.4.1 Initial Tax Microsimulation on un-modified Survey Data

The underlying database for this exercise is the Sakernas 2008 raw database. We preserved this as our unmodified base file. This is unmodified in the sense that we are not imputing information from other databases into it. Pre-adjustment and pre-conditioning of this unmodified data was still needed. Alignment to the national publication of average provincial labour salary, and allocation of the possible dependents to eligible tax payers from the income earners identified in each household to prepare the database for tax estimation were explained in detail in Chapter II of the thesis.

Table 19. Initial simulation of the un-modified Raw Sakernas 2008

<table>
<thead>
<tr>
<th>Deciles of Tax Payers</th>
<th>Average Annual Income (000)</th>
<th>Average No of Dependents</th>
<th>Average Non Taxable Income (000)(*)</th>
<th>Average Income Tax Liability (000)</th>
<th>Impact of Exemptions on reduction of tax liabilities (000)(***)</th>
<th>Dependents Impact on reduction of tax liabilities (000)(***)</th>
<th>Total Income Tax Liabilities (000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17,315.05</td>
<td>2.00</td>
<td>16,272.65</td>
<td>813.63</td>
<td>153.63</td>
<td>69,627,833</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19,130.65</td>
<td>2.58</td>
<td>17,204.31</td>
<td>860.22</td>
<td>200.22</td>
<td>128,737,448</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>20,960.08</td>
<td>2.64</td>
<td>17,321.72</td>
<td>866.09</td>
<td>206.09</td>
<td>243,070,636</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>22,848.96</td>
<td>2.63</td>
<td>17,340.27</td>
<td>867.01</td>
<td>207.01</td>
<td>368,098,166</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25,030.54</td>
<td>2.65</td>
<td>17,342.99</td>
<td>887.01</td>
<td>207.15</td>
<td>513,612,253</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>28,367.39</td>
<td>2.65</td>
<td>17,390.79</td>
<td>1,037.91</td>
<td>209.54</td>
<td>733,384,977</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>31,799.16</td>
<td>2.66</td>
<td>17,413.00</td>
<td>1,210.61</td>
<td>210.65</td>
<td>961,174,018</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>36,756.67</td>
<td>2.65</td>
<td>17,414.78</td>
<td>1,459.88</td>
<td>218.25</td>
<td>1,290,600,307</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>43,760.09</td>
<td>2.63</td>
<td>17,422.33</td>
<td>1,704.64</td>
<td>384.64</td>
<td>1,898,814,164</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>95,737.52</td>
<td>2.65</td>
<td>17,429.25</td>
<td>2,784.06</td>
<td>631.22</td>
<td>15,016,308,389</td>
<td></td>
</tr>
</tbody>
</table>

(*) = Non Taxable Income/Number of Household Member
(**) = Tax Liabilities based on Gross Income - Normal Tax Liabilities
(***)= Tax Liabilities when dependent assumed zero-Normal Tax Liabilities

We will also assess the representativeness of several alternative base files with our tax microsimulator. The representativeness test was done by comparing the result from each database to the figure of the national collection of personal income tax in 2008 as the benchmark. Simultaneously we analysed the interactions between several components in the
calculation of income tax liabilities with each alternative database. Later we will see the interactions between the different components of personal income tax calculations from the different deciles in the different databases.

The main characteristic of the family which influences the calculation of taxable income is the number of dependents. It is shown in the third column of Table 19. We assumed full take-up benefits for this simulation meaning all tax payers were assumed to maximise their eligible personal exemptions by claiming the maximum number of dependents. Surprisingly, we found that the lowest decile households by Sakernas sampling had the lowest number of average dependents (two dependents), while the other deciles averaged 2.6. This may be due to tax planning, where higher income earners will take more dependents into their family. The impact of this will be reduced in this model because of the assumption that incomes are combined to the household level.

It may also be that lower income households have a lower number of household members so the representative sampling from the lowest decile of tax payers income in Sakernas will benefit households with relatively lower personal exemptions compared to households in higher income deciles.

We would like to see how significant the reduction in income tax liabilities is from the application of personal tax exemptions. In this Indonesian PIT case, the tax exemption is represented by non-taxable income entitlements as stipulated in the tax law 17 (2000). We analysed this by comparing income tax liabilities calculated on purely gross income (without taking any exemptions out) with the income tax liabilities calculated based on taxable income (gross income minus the eligible exemptions). The reduction in the income tax liabilities
from the previous comparison is our proxy for the impact of tax exemptions as shown in column 6 of Table 19.

The effect of dependents on the income tax liabilities is shown in column 7 of Table 19. The dependent effect was simulated firstly by recalculating income tax liabilities from all of the taxpayers assuming they have zero dependents. After that we then deducted the actual income tax liabilities which were calculated based on their actual dependent’s entitlements.

From column 6 of Table 19, we noted that the impact of the application of eligible tax exemptions on the reduction of the income tax liabilities in every decile are somehow bigger than the income tax liabilities burden per decile of taxpayers. Taxpayers in deciles one to three experienced lower annual income tax liabilities (52 thousands rupiah to 182 thousands rupiah) compared to the tax reduction impact on the dependent (153 thousands rupiah in decile one to 206 thousands rupiahs in decile three). Deciles four to ten have higher income tax liabilities than the tax reduction impact on the dependent. It seems that the first, second and third deciles experience a more favourable impact if they fully maximise their tax exemption entitlements by claiming the maximum number of dependents. The non-taxable entitlement is one of the crucial instruments in the Indonesian PIT system which is regularly amended to adjust for inflation and the basic standard of living of the Indonesian household. We will explain the rulings and reforms of this component in Chapter IV. Contrary to our favourable findings for deciles one to three, the result from column 7 Table 19 about the proportion of the dependents effect, shows decile ten benefits by a more than four-fold reduction in annual income tax liabilities (631 thousand rupiah) compared to the reduction in annual income tax liabilities of the first decile tax payers (153 thousand rupiah).
To sum up the result, although the application of our tax microsimulator to the original/unmodified database has illustrated a different distribution of tax paid for different deciles of income, the calculation shows the lack of coverage of high income earners in the Sakernas database. One of the results from applying the tax simulator to the un-modified data of Sakernas was a high concentration of income tax liabilities. Almost 71 percent of the income tax is contributed by the highest income earners in decile ten. In addition to this high concentration, the end result of the total tax liabilities indicated a significant underestimation of the income tax from labour/employment. The percentage of simulated revenue was only 41 percent from the actual realised labour income tax revenue in 2008. This underestimation could be a function of other income, however given time and data limitations this thesis only focuses on labour income and its underestimation.

This strengthens the argument in Chapter II that the major issue in the estimation of the lower and higher incomes in Indonesia is the unavailability of good income data. This has been an ongoing problem in the study of Indonesian income distribution (Cameron, 2002; Leigh & van der Eng, 2009). Therefore, this brings us to the use of the next two databases.

### 3.4.2 Initial Tax Microsimulation on modified Survey Data

**Susenas Imputed with Sakernas Labour Income**

As previously explained we emphasised the use of Susenas as the main option for the Indonesian income tax base data due to its long history, richer household variables, bigger sample and wider area coverage. Susenas has become the major source of data for Indonesian
income based-studies as it is the oldest regular national survey collecting Indonesian social and economic indicators.

This Susenas-modified database is a result of our imputation process from the previous exercise which is already explained in detail in Chapter II. It contained full characteristics of Susenas’ labour income earners adding previously missing labour income information by imputing the information using Sakernas labour income information. The imputation exercise uses a similar person in the Susenas database to impute the information on labour income from the Sakernas database using demographic, geographic and socio-economic characteristics.

Table 20. Initial simulation of the Susenas imputed with Sakernas Labour Income 2008

<table>
<thead>
<tr>
<th>Deciles of Tax Payers</th>
<th>Average Annual Income (000)</th>
<th>Average No of Dependents</th>
<th>Average Non Taxable Income (000)(*)</th>
<th>Average Income Tax Liability (000)</th>
<th>Impact of Exemptions on reduction of tax liabilities (000)(**)</th>
<th>Dependents Impact on reduction of tax liabilities (000)(***)</th>
<th>Total Income Tax Liabilities (000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14,603.05</td>
<td>0.04</td>
<td>13,541.73</td>
<td>53.07</td>
<td>677.09</td>
<td>17.09</td>
<td>76,762,475</td>
</tr>
<tr>
<td>2</td>
<td>17,094.91</td>
<td>0.70</td>
<td>14,764.09</td>
<td>116.54</td>
<td>738.20</td>
<td>78.20</td>
<td>168,582,524</td>
</tr>
<tr>
<td>3</td>
<td>18,929.38</td>
<td>1.58</td>
<td>16,001.67</td>
<td>146.39</td>
<td>800.08</td>
<td>140.08</td>
<td>211,856,022</td>
</tr>
<tr>
<td>4</td>
<td>20,694.23</td>
<td>1.59</td>
<td>16,039.09</td>
<td>232.76</td>
<td>801.95</td>
<td>141.95</td>
<td>336,729,252</td>
</tr>
<tr>
<td>5</td>
<td>23,022.71</td>
<td>1.64</td>
<td>16,114.11</td>
<td>345.43</td>
<td>805.71</td>
<td>145.71</td>
<td>499,786,775</td>
</tr>
<tr>
<td>6</td>
<td>25,739.14</td>
<td>1.66</td>
<td>16,144.87</td>
<td>479.71</td>
<td>850.57</td>
<td>147.24</td>
<td>694,039,332</td>
</tr>
<tr>
<td>7</td>
<td>29,399.55</td>
<td>1.65</td>
<td>16,158.06</td>
<td>662.07</td>
<td>1,027.88</td>
<td>147.90</td>
<td>957,912,935</td>
</tr>
<tr>
<td>8</td>
<td>34,036.72</td>
<td>1.76</td>
<td>16,319.94</td>
<td>885.84</td>
<td>1,267.83</td>
<td>156.00</td>
<td>1,281,328,581</td>
</tr>
<tr>
<td>9</td>
<td>40,881.71</td>
<td>1.83</td>
<td>16,429.63</td>
<td>1,273.95</td>
<td>1,564.22</td>
<td>246.75</td>
<td>1,843,452,266</td>
</tr>
<tr>
<td>10</td>
<td>85,355.48</td>
<td>1.79</td>
<td>16,347.00</td>
<td>9,314.79</td>
<td>2,412.06</td>
<td>436.15</td>
<td>13,478,546,571</td>
</tr>
</tbody>
</table>

(*) = Non Taxable Income/Number of Household Member

(**) = Tax Liabilities based on Gross Income - Normal Tax Liabilities

(***)= Tax Liabilities when dependent assumed zero-Normal Tax Liabilities

The total income tax liabilities produced by our tax simulator using the Susenas imputation database shown in table 20 is slightly lower than the original source of labour-income information (Sakernas 2008). A decrease of 3 percent (the percentage of PIT simulated result
from the actual revenue of 38 percent compared to 41 percent from the result of previous database) was noticed as a result of this estimated tax from the imputed base file. So far, of the two imputations onto our base file, this is the best result we could get. It is not too far from the original sources but there is a slight indication of lower coverage of high income earners (average gross income in decile ten where the imputed income was 85 million rupiah) compared to the source database (95.7 million rupiah). However, even with slightly lower coverage of higher income earners, we still note that the contribution from the highest income decile compared to the total income tax liabilities is the same (about 69 percent from the total liabilities). This result preserved the dependency of the current Indonesian PIT system on highest income decile for the biggest contributor to the PIT national income.

**Susenas Imputed with Sakernas Labour Income and PIT Returns**

The final database is a combination of reweighted PIT return micro data and the Susenas imputed Sakernas database. The main aim of this imputation is to increase the coverage of higher income earners as explained in previous chapter. We started the process by identifying linking variables to match PIT with the Susenas imputed Sakernas database. We then regrouped the database of income earners from PIT administrative systems to seventeen income brackets as described in Chapter II. The end result of this exercise was an additional 2.2 million sample taxpayers which could be added to the 69,013 eligible tax payers which were previously extracted from 274,390 income earners from the Sakernas’ donor database. Therefore we might expect a totally different result from the analysis compared to our previous unmodified and imputed databases.
Table 21. Initial simulation of the modified Susenas imputed with Sakernas and Tax Return 2008

<table>
<thead>
<tr>
<th>Deciles of Tax Payers</th>
<th>Average Annual Income (000)</th>
<th>Average No of Dependents</th>
<th>Average Non Taxable Income (000)(*)</th>
<th>Average Income Tax Liability (000)</th>
<th>Impact of Exemptions on reduction of tax liabilities (000)(**)</th>
<th>Dependents Impact on reduction of tax liabilities (000)(***)</th>
<th>Total Income Tax Liabilities (000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14,653.03</td>
<td>0.05</td>
<td>13,530.57</td>
<td>56.12</td>
<td>676.53</td>
<td>16.53</td>
<td>92,997,078</td>
</tr>
<tr>
<td>2</td>
<td>17,102.50</td>
<td>1.11</td>
<td>15,027.82</td>
<td>103.73</td>
<td>751.39</td>
<td>91.39</td>
<td>171,803,427</td>
</tr>
<tr>
<td>3</td>
<td>18,825.66</td>
<td>1.53</td>
<td>15,462.72</td>
<td>168.15</td>
<td>773.14</td>
<td>113.14</td>
<td>278,788,027</td>
</tr>
<tr>
<td>4</td>
<td>20,755.12</td>
<td>1.51</td>
<td>15,458.73</td>
<td>264.82</td>
<td>772.94</td>
<td>112.94</td>
<td>438,796,057</td>
</tr>
<tr>
<td>5</td>
<td>23,101.86</td>
<td>1.57</td>
<td>15,541.97</td>
<td>388.44</td>
<td>777.10</td>
<td>117.10</td>
<td>643,812,469</td>
</tr>
<tr>
<td>6</td>
<td>26,409.56</td>
<td>1.51</td>
<td>15,437.73</td>
<td>548.59</td>
<td>843.84</td>
<td>111.89</td>
<td>909,073,052</td>
</tr>
<tr>
<td>7</td>
<td>30,379.35</td>
<td>1.52</td>
<td>15,454.27</td>
<td>746.25</td>
<td>1,041.68</td>
<td>112.71</td>
<td>1,236,481,295</td>
</tr>
<tr>
<td>8</td>
<td>35,745.29</td>
<td>1.69</td>
<td>15,680.47</td>
<td>1,003.59</td>
<td>1,320.94</td>
<td>126.98</td>
<td>1,663,213,786</td>
</tr>
<tr>
<td>9</td>
<td>43,993.09</td>
<td>1.63</td>
<td>15,568.29</td>
<td>1,608.32</td>
<td>1,543.98</td>
<td>220.99</td>
<td>2,665,275,666</td>
</tr>
<tr>
<td>10</td>
<td>153,257.40</td>
<td>1.40</td>
<td>15,052.32</td>
<td>28,737.12</td>
<td>2,842.74</td>
<td>324.35</td>
<td>58,584,253,912</td>
</tr>
</tbody>
</table>

(*) = Non Taxable Income/Number of Household Member
(**) = Tax Liabilities based on Gross Income - Normal Tax Liabilities
(***)= Tax Liabilities when dependent assumed zero-Normal Tax Liabilities

Sufficient coverage of higher income earners leads to a better estimation of the income tax from labour or employment activities due to the addition of 2.2 million tax payers which is distributed across every decile of income earners and has increased the figure of total income tax liabilities significantly. If we compare the result with the previous database, the increase in the total tax liabilities is experienced by every decile as shown in table 21. The increase for decile one is 21 percent while decile two experienced the lowest increase of only 1.9 percent. The increase from in total tax liabilities from decile three until decile eight is around 30 percent on average. The more significant increase in total tax liabilities of 44.5 percent was seen in decile nine. In the previous Susenas imputation database this decile generated total tax liabilities of 1.8 trillion rupiah, and this has increased to produce total income tax liabilities of 2.6 trillion rupiah due to the additional taxpayers from the PIT imputation.

Decile ten shows the most contribution to the PIT datasets with an increase of more than three fold. It started with 13.5 trillion in the previous Susenas-Sakernas imputation and increased to 58.6 trillion rupiah with the current Susenas-Sakernas-PIT imputation. In line with the significant increase of decile ten’s total tax liabilities, we also found an 87.85
percent total contribution of tax liabilities. This indicates a far higher concentration of high income earners compared to the previous databases. In addition to that the 29 percent higher proportion of the simulated revenue compared to the actual revenue in 2008 should confirm that not all Indonesian income earners were paying income tax. This finding is in line with what Tanzi and Zee (2001) observed as one of the main challenges to develop better and more efficient tax structures for developing countries: the size of its informal economy. In the Indonesian case, Bureau of Statistics Indonesia (BPS) estimates that in 2006, informal employment was about 64 percent and the share of small and mostly informal sector’s enterprises to the GDP output was roughly 38 percent (BPS, 2010).

3.4.3 What we learned from tax microsimulation

Comparison of the modified and unmodified databases

By applying the income tax simulator to the different databases, we could analyse the importance of each step of the data modification. By only applying the tax simulator to the unmodified survey micro data we can already extract much information on the characteristics of taxpayers disaggregated by decile of income. However, the application of the tax simulator to the original database shows how the data base is lacking observations of high income earners and is unable to simulate its potential total tax liabilities that can be achieved by using the imputed databases.
As indicated by the previous findings of Ikhsan, Trialdi & Syahrial (2005) and Leigh & van der Eng (2009) on the concentration of the highest income decile tax payers, the result of applying the tax simulator to the un-modified data of Sakernas confirmed the high concentration of income tax liabilities. Almost 71 percent of the income tax is contributed by the highest income earners in decile ten. In addition to this high concentration, the end result of the total tax liabilities indicated a significant underestimation of the income tax from labour/employment. This also indicated a problem of low observations of high income earners from the Sakernas database.

That indication might lead to our findings of the average effective tax rate from our simulation on the unmodified data as shown in table 22. The average effective tax rate calculated in table 22 is the average income tax liability (in column 5 of Table 19 or Table 20) divided by the average gross income (in column 2 of Table 19 or table 20). By seeing the result in table 22, our tax simulator had shown us that the average effective tax rate from each of the deciles is very low; far lower than the statutory tax rate.
From table 22, it is shown the average effective tax rate from the source of the labour income information, the Sakernas raw data, ranged from 0.3 percent in decile one then increases gradually to more than 1.2 percent in decile four up until decile seven with more than 2.2 percent. Decile nine showed another jump to 3.2 percent while in the highest decile ten, the average effective tax rate was 7.4 percent. Further, if we compare the information from the average effective tax rate, all deciles show a relatively close result. The average effective tax rate per decile of the imputed database was less than 0.1 percent different from its Sakernas’ source. There were only three deciles (deciles two, nine and ten) with an average effective tax rate difference of more than 0.1 percent. The highest relative difference noted in decile ten is 0.59 percent between the Sakernas’s and Susenas imputation’s average effective tax rate.
The problem of the lack of observations of the higher income earners was preserved both in the unmodified and modified base data. The findings are the same from the unmodified and modified database: only decile ten produce an average effective tax rate that was higher than the lowest statutory tax rate of the Indonesian personal income tax (PIT) system based on Law 17(2000) which is 5 percent for the first income bracket after the deduction of non-taxable income. This gradual increase of average effective tax rate showed the tax progression from each decile of tax payers, however it is very different from the statutory rate which ranges from 5 percent to 35 percent.

Those findings were in line with the distribution of the income group category based on law 17(2000) and also the distribution of maximum gross income from each decile in both databases. Both the unmodified Sakernas and modified Susenas database showed that the majority of income earners are low income earners with most of the taxable income from the decile categorised as income group one (with taxable income of less than 25 million rupiahs or 2,500 Australian dollars annually). The Sakernas source covered more high income earners and the move to income group two started in decile 8 of the tax payers’ gross income, whereas Susenas modified only started to move to income group two in decile 9. Income group two consisted of taxpayers with taxable income between 25 million rupiahs and 50 million rupiahs annually.
The Contribution of Higher Income Earners from the Tax Administrative Data in the Modified Database

Almost 70 percent of the tax liabilities were contributed by the imputed tax payers from the tax return data. This should answer the ongoing issue of under coverage of high income earners found in several Indonesian income-based studies (Cameron, 2002; Leigh & van der Eng, 2009) and which leads to a better income micro-data to better understand the structure of Indonesian income and the impact of government policy intervention. This will be elaborated on more by studying the impact of the latest PIT reform in Chapter IV.

However, as we have explained previously in Chapter II, the tax return administrative data is a compilation of a low compliance tax return submission. In addition to that, we also observed some missing values and variables in our tax return data due to recording errors. However, we still got a better result from our choice to maximize the usefulness of the administrative data in providing the information for both the lower and the top income tax payers. By this we might expect a better and more accurate analysis of the Indonesian potential tax revenue simulation given the compliance in tax reporting is higher. The compliance rate in 2008 for the personal income tax payer is 31% (Directorate General of Taxes Indonesia, 2008).

3.5 The Sensitivity of the Simulator to Tax Payers’ Dependents

These simulations are being used to do sensitivity testing of our personal income tax microsimulator. Different characteristics of the sampled tax payers in the survey data will
impact the revenue and distribution of income tax liabilities differently. This change to the
different characteristics of tax payers can be simulated with our tax simulator. We would like
to test the usage of our tax simulator to know how different characteristics of households will
impact on the different deciles of the tax payers.

We will start our simulation by revisiting the basic formula of the Indonesian PIT system.
The taxable income is calculated by deducting the non-taxable income from the annual gross
income. Our simulator will calculate income tax liabilities from the taxable income. Once the
taxable income is calculated, the average effective tax rate per decile income of the tax payer
can be estimated from the data. Since the base data came from two different surveys, the
application of the tax simulator to Sakernas and Susenas based data will tend to produce
different results. Previously, as explained in chapter 2, by seeing the result from the
unmodified and modified database, we found a significant difference in the characteristics of
income earners in the Susenas’ imputation. It shows the importance of using the database
with more detailed information on the number of children and the number of dependents in
the household to show the true extent of the reduction. The Susenas 2008 base file, referring
to table 20 and table 21, has a lower average number of dependents per tax payers compared
to Sakernas 2008 (referring to table 19), which has a higher average number of dependents
per decile of taxpayers’ income. It means that due to the different data sources, the number of
household members in households with income is lower in Sakernas than Susenas.
Table 23. Comparison of dependent distribution Sakernas 2008 and Susenas 2008 imputation

<table>
<thead>
<tr>
<th>No of Dependents</th>
<th>Sakernas Raw Weighted Observations</th>
<th>Susenas Imputed with Sakernas Weighted Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>626,290</td>
<td>5,874,577</td>
</tr>
<tr>
<td>1</td>
<td>810,687</td>
<td>1,477,017</td>
</tr>
<tr>
<td>2</td>
<td>2,179,730</td>
<td>2,234,898</td>
</tr>
<tr>
<td>3 &amp; more</td>
<td>9,746,072</td>
<td>4,881,515</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>13,362,779</strong></td>
<td><strong>14,468,007</strong></td>
</tr>
</tbody>
</table>

Table 23 shows that Sakernas has a higher concentration of tax payers with three dependents while Susenas has a higher concentration of the tax payers with zero and three dependents.

Table 23 shows the average number of dependents per decile. The previous results from the application of tax exemptions and the impact of dependents on the reduction of the income tax liabilities to the Susenas-Sakernas-PIT imputation data as shown in table 21 shows a similar pattern to the previous Susenas-Sakernas imputation showed in table 20. In sum, the higher the decile the higher will be the reduction of tax liabilities from tax exemptions and dependents.

Table 24 shows the impact of the different dependent simulation. To conduct this simulation, firstly we set up a characteristic of the taxpayer’s household to four sets of conditions. The first assumed all taxpayers have no dependent (so they are only entitled to claim their own status of employment in their tax exemption). The second assumed all taxpayers have one dependent, the third assumed all taxpayers have two dependents, and the last one assumed all taxpayers have three dependents. Table 24 below is the summary assessment of the impact of different dependents on different income deciles. We focused our impact analysis on the average effective tax rate for each decile. It is an approximation of the average tax burden on tax payer’s income for each decile.
The calculation of average effective tax rate (Table 24) explains that claiming more dependents will decrease the average tax burden of the taxpayers. The decreased burden is presented by the decrease in the average effective tax rates. The average effective tax rate was a decile’s weighted average of income tax liability divided by gross income. Hence those results were derived from the interaction of the dependent entitlement, marital status, income groups, the applicable progressive income tax rates and a weighted sample from the base file. In line with the ideal conditions, our microsimulation results confirmed that claiming more dependents (assuming full take up of the entitlement) will reduce the burden on the taxpayers. The movement from lower to higher decile also confirmed that the average effective tax rate from the microsimulation model resulted in a consistent increase with all four set up conditions. The most significant changes in the average effective tax rates were shown by the result for decile ten by a jump from one digit of percentage to two digit. It confirmed that the concentration of higher income earners (with annual taxable income more than 200 million) also happened in decile 10.

### Table 24. Different Dependent Simulation Susenas-Sakernas-PIT Imputed base file by decile

<table>
<thead>
<tr>
<th>Deciles of Tax Payers</th>
<th>Average Effective Tax Rate</th>
<th>Dependent 0</th>
<th>Dependent 1</th>
<th>Dependent 2</th>
<th>Dependent 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.18%</td>
<td>-0.23%</td>
<td>-0.65%</td>
<td>-1.06%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.56%</td>
<td>0.18%</td>
<td>-0.20%</td>
<td>-0.59%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.99%</td>
<td>0.65%</td>
<td>0.30%</td>
<td>-0.04%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.39%</td>
<td>1.08%</td>
<td>0.77%</td>
<td>0.46%</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.79%</td>
<td>1.52%</td>
<td>1.24%</td>
<td>0.97%</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2.20%</td>
<td>1.96%</td>
<td>1.72%</td>
<td>1.48%</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2.60%</td>
<td>2.40%</td>
<td>2.19%</td>
<td>1.99%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2.98%</td>
<td>2.81%</td>
<td>2.64%</td>
<td>2.47%</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3.99%</td>
<td>3.74%</td>
<td>3.50%</td>
<td>3.29%</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10.58%</td>
<td>10.37%</td>
<td>10.17%</td>
<td>9.98%</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>6.78%</strong></td>
<td><strong>6.55%</strong></td>
<td><strong>6.32%</strong></td>
<td><strong>6.09%</strong></td>
<td></td>
</tr>
</tbody>
</table>
What has been shown is that claiming one dependent led to a decrease in the tax burden relative to claiming no dependent, and claiming three dependents yields a lower burden compared to claiming two dependents. In sum we might expect tax payers who claim more dependents will pay less income tax compared to those who claim fewer dependents. This sensitivity analysis on the microsimulation model has provided important information on the impact of the current application of these lump-sum deductions on tax payers’ dependents from the Indonesian PIT law.

3.6 Application for Future Policy Simulation

The major benefit of having a static tax simulator is its ability to assess the distributional impact of the gains or losses for each micro-unit from future policy. There are several variations of the rules of PIT eligibility that we have simulated. The impacts can also be assessed for specific small units of analysis (for example, within the certain threshold of low income earner in particular municipalities or within the tax payer with two income earners and three children in the household) or further aggregated units of analysis (national populations). Hence, the microsimulation model enables users to develop policy parameters to adapt government policy objectives and to assess the distribution impact and to determine the winners and losers from the application of policy (Citro & Hanushek, 1991; Gupta & Harding, 2007).

The other benefit of static microsimulation models is that they provide a very detailed disaggregation of the revenue change itself (for revenue estimation). Even though not dynamic
in the supply-side sense, microsimulation is held as the standard in many tax authorities for producing revenue estimates due in part to controversy over dynamic assumptions.

3.6.1 The Benefit from Applying Different Maximum Dependent for Tax Relief Component

There are likely several policies to be pursued by the Indonesian government in the future. An assessment of the law 36 (2008) as the latest PIT reform, and its non-taxable income component will be considered in detail in Chapter IV. Here we aim to endorse the power of the tax simulator by simulating a hypothetical policy option. The option will be to adjust the number of dependent entitlements to the distribution of dependents based on the modified Susenas base data. As shown in Table 25, the majority of the tax payers in the modified database claimed zero dependents or three dependents. As an addition to the analysis in Table 24 which assumes all tax payers have the same number of dependents, we would like to extend our hypothetical future analysis to look at what would happen if the stipulated PIT law included more or fewer dependents in their tax relief component.

Table 25 below shows the impact on revenue and distribution of burden for four scenarios. Provided that all other rulings stay the same (ceteris paribus), we will only change the maximum number of dependents entitlement. So scenario 1 is based on the current rulings (the maximum number of dependents is three), scenario 2 assumes the maximum of dependent is two, scenario 3 assumes the maximum is one, and the last scenario assumes no dependents can be claimed.
In terms of its total impact, from Table 25 we found that by applying different maximum dependent entitlements, the government will consistently yield more revenue if the stipulated tax law allows for fewer dependents in the tax relief component. In conjunction with that, we found a similar impact on the increasing revenue impact; the stipulation of a smaller dependent entitlement will increase the burden on the tax payers in all the cases. The total highest tax burden, or the national average effective tax rate 6.72%, resulted from the stipulation of a maximum of nil dependents which could be claimed. There were various results from various deciles but the total, or national, tax burden results suggest that application of a nil dependent entitlements will have the highest burden on tax payers. The second highest impact results from the application of a maximum of one dependent which
results in a very small difference compared to the two dependents entitlement and the lowest burden achieved is by the stipulation of three dependents entitlement.

### 3.6.2 The Benefit from Different Taxable Income Threshold

In the second policy simulation, we simulated the impact from the application of the increasing taxable income threshold stipulated by the law. We started with an increase of 1 million rupiahs for each of the taxable income thresholds, then an increase of 2 million rupiahs, of 5 million rupiahs, and lastly an increase of 10 million rupiahs. All other stipulated regulations were assumed to remain the same. The result of this likely scenario is in Table 26 below.

#### Table 26. Revenue and Distribution of Burden from Different Taxable Income Threshold

<table>
<thead>
<tr>
<th>Deciles of Tax Payers</th>
<th>Potential Revenue Impact (Total Income Tax Liabilities in Billion Rupiahs)</th>
<th>Distribution of Burden (Average Effective Tax Rates)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income Increase by 1 million</td>
<td>Income Increase by 2 million</td>
</tr>
<tr>
<td>1</td>
<td>47</td>
<td>47</td>
</tr>
<tr>
<td>2</td>
<td>153</td>
<td>153</td>
</tr>
<tr>
<td>3</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td>4</td>
<td>463</td>
<td>462</td>
</tr>
<tr>
<td>5</td>
<td>673</td>
<td>673</td>
</tr>
<tr>
<td>6</td>
<td>949</td>
<td>949</td>
</tr>
<tr>
<td>7</td>
<td>1,290</td>
<td>1,290</td>
</tr>
<tr>
<td>8</td>
<td>1,750</td>
<td>1,750</td>
</tr>
<tr>
<td>9</td>
<td>3,090</td>
<td>3,030</td>
</tr>
<tr>
<td>10</td>
<td>57,100</td>
<td>56,800</td>
</tr>
<tr>
<td><strong>TOTAL REVENUE AND AVERAGE EFFECTIVE TAX RATES</strong></td>
<td>65,809</td>
<td>65,448</td>
</tr>
</tbody>
</table>
Total impacts on the revenue and distribution of the tax burden are similar, meaning that the application of higher taxable income thresholds will decrease the government revenue. In line with the decreasing the revenue yield, the tax burden of the taxpayers is also decreased.

However, an interesting impact to be assessed is the differential impact in deciles nine and ten. For decile one up to decile eight, the impact of the revenue yield and the average effective tax rates is similar. Consistently decreasing revenue and a decreasing tax burden were found, starting with decile nine. This might be associated with the previous findings about the concentration of the higher income earners in decile nine and ten. Only the decile with wealthy income earners benefits from the policy to increase taxable income thresholds.

3.7 Conclusion

The strength of the microsimulation model lies in its ability to include whole characteristics of the economic agent within the heterogeneous sample of the individual micro-data and to give a detailed overview of the tax reforms’ distributional impact on them. After the individual impacts were captured, then the financial cost and benefits of the reforms as an aggregate were easy to calculate (Mitton, Sutherland & Weeks, 2000; Bourguignon & Spadaro, 2006).

In this chapter we developed a more detailed microsimulation for Indonesian personal taxpayers and focused the scope of the second key factors on building labour income tax microsimulation by coding the rules of labour income tax from income tax law no 17 year
2000. This followed from Chapter II which enhanced the first key factor of microsimulation, the micro-dataset of information on the economic agents (individuals and/or households).

We used administrative tax returns as the main benchmark for the variables to build a tax simulator for the Indonesian labour-income tax microsimulation model. Two groups of benchmark variables were derived from the personal income tax returns; individual information and income detail for tax calculation. We identified the similar variables provided by both our Sakernas donor survey data and Susenas cold deck imputed survey data for use in the tax simulator. As argued by Atkinson (1997) and Paas (1986), we adjusted the available variables in the survey for use in our tax simulator. The results of our transposition of households to individual units of observations, determination of the dependent for applying eligible exemptions, grossing up and aligning the income unit, lead us to build our preliminary assumptions for the tax microsimulation model. Once the built-in assumptions and the main income dataset were completed, we built our parameter dataset for our tax simulator by following Yuwono (2008) with some modifications.

We then applied our microsimulation model to several different base data to weigh up the benefits and limitations of using the raw or unmodified database compared to the modified ones. We concentrated our simulations on the deciles of tax payers. We calculated the average income, average number of dependents, average non-taxable income, and average income tax liabilities. From each of the decile’ calculations we proceeded to simulate the impact of total tax reduction from the different taxable incomes by comparing income tax based on gross income and income tax based on taxable income. Finally we simulated the
total tax reduction experienced by different tax payers with different household characteristics (in this case represented by their different number of household members).

Several important results from our test on the usage of our tax microsimulation model can be summarized as follows:

1. The calculation of our microsimulation model using the unmodified database shows a lack of coverage of high income earners from the Sakernas. Another obvious result is the high concentration of income tax liabilities and a significant underestimation of the income tax from labour/employment.

2. The response of our tax microsimulation model to the first modified database, an imputed Sakernas income to Susenas database (which previously has a total missing income) yields a slightly lower tax estimation (but the difference is not too far from the sources). We also noted that the result of this first imputation preserved the dependency of the current Indonesian PIT system on the highest income decile. The highest income decile (decile ten of the tax payer’s income) is still the biggest contributor to the PIT national income.

3. A better result was achieved by the application of the model to the latest modified database. By imputing higher income earners from PIT returns to every decile of the Susenas imputed database, sufficient coverage of higher income earners could be achieved. There are an additional 2.2 million of them scattered across almost every decile of income earners. This increased the figure of total income tax liabilities significantly (70% of tax liabilities is contributed by the imputation of tax payers from tax return data). Hence, it leads to a better estimation of the income tax from labour or employment activities.
4. Both simulations of the modified and unmodified data base yield a low average effective tax rate. In most deciles we found that the average effective tax rate is far lower than the statutory tax rate stipulated by the PIT Law. It is only in the highest decile (decile ten) the tax payers bear the burden slightly higher than 10% or slightly higher than the second income bracket’s statutory rate.

5. We can use our tax estimator to assess how different conditions will impact on different deciles of tax payers. Our assessment of how changes in the application of tax exemptions and in the number of dependents impact on the reduction of the income tax liabilities showed that the result from the Susenas-Sakernas-PIT imputation data has a similar pattern to the results from Susenas-Sakernas imputation. In sum, the higher the decile, the higher will be the reduction of tax liabilities because of tax exemptions and dependent’s applications.

6. We also confirmed the result of the average tax burden from the same dependent condition from the exercise of our tax microsimulator on the Susenas-Sakernas-PIT imputed data. We simulated four sets of conditions: assuming all taxpayers in all deciles have zero dependents, one dependent, two dependents and three dependents. Ideally we should expect that claiming more dependents (assuming full take up of the entitlement) will reduce the burden of the taxpayers. Our results confirmed the expectation, and table 24 shows that claiming more dependents will decrease the average tax burden of the taxpayers.

7. We then simulated the impact of the application of the different maximum dependent-based exemptions on tax payers. In terms of its total impact, we found that the government will consistently yield more revenue if the stipulated tax law allows smaller dependent entitlements in the tax relief component. We also found a supporting
impact on the revenue; the stipulation of less dependent entitlement will increase the burden of the tax payers in general.

8. Our last policy simulation assessed the impact from the application of increases in the taxable income threshold. The simulation started with a 1 million rupiah increase of each of the taxable income thresholds, then a 2 million rupiah increase, a 5 million rupiah increase and lastly an increase of 10 million rupiah. Interestingly, the impact on the revenue yield and the average effective tax rates is similar for most deciles (decile one up to decile eight). It only appears that the real impact starts to happen for tax payers in deciles nine and ten. The application of higher taxable income thresholds reduced the government revenue and in line with the decreasing yield, the tax burden of the taxpayers also decreased.
CHAPTER IV. REVENUE AND DISTRIBUTION IMPACT ANALYSIS OF
INDONESIAN PERSONAL INCOME TAX REFORM IN 2008

4.1 Introduction

As mentioned in chapter I, tax has become increasingly important for the Indonesian economy since the 1980’s. The Indonesian Government has initiated a fundamental reform of the tax system and start reforming the income tax law to save considerably reduced Government revenue. The plunge in the revenue at the same time as an increasing proportion of foreign debt drove the need to increase revenue (Anwar, Azis, Pangestu & Soesastro, 1991; Hill, 2000; Resosudarmo & Kuncoro, 2006). One of the major reforms was to apply a new self-assessment system. This reform replaces the official-assessment system that had been implemented since the Dutch colonial era. At the same time, the reform also simplifies both the personal and corporation tax rate that allowed a lower rate and broader tax base to be enforced with better tax administration systems (Heij, 2000).

The 1983 tax reform was successful in increasing tax compliance and becoming the key element to support the Indonesian government budget and to maintain Indonesian fiscal sustainability after the oil crisis (Ikhsan, Tialdi & Syahrial, 2005). After the 1983 tax reform, especially since the 1990s, further reforms of various types of taxes continued to aim for simpler, fairer and more neutral tax laws. Simplicity and fairness was achieved by reducing and simplifying the income brackets and the rates progressivity, while neutral tax laws were achieved from the application of global rather than scheduler systems. It meant that all of the
income was pooled into one basket of income for the calculation of the income tax. These gradual reforms provided smooth shifting from the old official-assessment system to the new self-assessment system (Directorate General of Taxes (DGT), 2007).

The latest of a string of tax reforms is the provision of Law no 36(2008). Despite maintaining personal income progression as the equalizing tool to improve tax neutrality and fairness among Indonesian income earners, this new tax law seems to reduce the tax burden through lowering the top income tax rate, increasing the lowest income brackets with a 5 percent tax rate and increasing non-taxable income. Tanzi & Zee (2001) strongly argued that the improvement of effective rate progression could be achieved by a reduction in the nominal rate progressivity, changing income brackets and reducing exemptions and deductions.

This study examines the provision of Law no 36(2008) as the latest reform of personal income tax by looking at these possibly conflicting objectives. Specifically, the research questions that will be answered are about the impact of the 2008 income tax reform on revenue and income distribution and the role of compliance in fulfilling the aim of the reform.

These questions will be answered by constructing the base data of Indonesian true income approximation in chapter II. We utilize the new dataset to capture full distribution of the Indonesian income (see chapter II). We will also utilize the detailed characteristics of national labour force survey (Sakernas) and national socio economic survey’s (Susenas) in the tax microsimulation tool described in chapter III. By doing so, it is possible to identify the
eligible taxpayer group from all salary earners in the household and, finally, examine the change in government tax collection across income groups and the distribution of the income tax burden.

Following this introduction, the second section will give an overview of the history of taxation and its reform in Indonesia. Section three will discuss PIT in Indonesia including its revenue performance and contribution to the overall tax revenue. This is followed by a fourth section about the empirical framework used in this study. Section five presents and discusses the result of the empirical estimation and section six concludes the study.

4.1.1 Aim of tax in Indonesia

From Heij (1993) and Uppal (2003), the motivation for Indonesia’s first radical tax reform in 1983 to the 2008 reform are:

1. Increased Non-Oil and Gas Tax Revenue/Revenue Adequacy

As Uppal (2003) argued, the Indonesian income tax structure had long been based solely on the colonial PIT and CIT Law. As discussed previously, all of these old laws put their basic focus on how to provide adequate cash flow to the colonial government. The first tax reform to remove these inefficient and outmoded colonial tax laws and administrations, and also to replace them with the Indonesian-own laws and tax systems, started in 1983. This new unified PIT and CIT income tax law began to be applied effectively in January 1984 with
performance of non-oil tax revenue (NOTR) being improved and increased significantly. The government’s dependency on oil-tax revenue could finally be reduced significantly and this is shown by the increasing proportion of government expenditure financed from the non-oil tax revenue (as what is shown in table 27). The series over the 1980’s to the 1990’s revealed that the money contribution from non-oil tax revenue increased by more than 25%. It clearly means that the government relied more and more on NOTR to finance its national budget. Unfortunately over the years the ongoing low tax to GDP ratios was still noted in the Indonesian tax structures. It implied an ongoing and relatively low coverage ratio from the Indonesian tax capacity. Indonesia somehow suffered from an inadequate revenue stream from its tax-collection effort. After the first reform was taken, there always some more parts of the Indonesian tax laws, systems and administration which could be improved.

2. Non Compliance

Again, Heij (1993) discussed that after the first radical reform was performed successfully, the administrative data from 1983 until 1990 showed significantly lower Indonesian tax collection and compliance compared to the neighbouring countries. With the population of around 170 million in July 1991, the registered tax payers was only 684,872 (0.4%) and only half of the registered tax payers actually paid income tax in 1991. From this, 22% of registered tax payers did not file their return. Uppal (2003) concluded that the main priority of tax administrative improvement and enforcement effort must be aimed at widening the tax base to reduce the ratio of non-compliance.
3. Economy

Low cost collection of taxes is one important indicator of an efficient tax system. As stipulated by the income tax law reform in 1983, the change from official-assessment to self-assessment should have significantly reduced the cost of tax collection. Along with that, the application of computer systems initially introduced to administer tax with the 1983 reform should boost the efficiency of the whole new income tax system. The efficiency of the system can be calculated using a cost of collection ratio measured using a ratio of how much the expenditure of the tax office (in order to perform their duty to collect tax revenue) compared to the realised revenue which is collected within the same year. The assessment could be both for aggregate figures or all tax offices (nation-wide) or figures from regional/sub-national level of the tax offices. Clearly it will show the efficiency of each of the offices in managing their available resources to collect revenue from the tax payers registered in its area of authority. The latest figures accessed from the DG’s 2008 Annual Report showed a relatively-efficient tax administration with cost of collection ratios to total tax revenue of 0.57 percent in 2007 and 0.56 percent in 2008, less than 1 percent from the national tax revenue collection. Looking at these figures, the Indonesian tax authority consistently maintained low cost collection ratios during these years, especially in comparison with its counterparts in the ASEAN countries.

4. Flexibility

Flexibility means that tax bases and rates should be set to follow the changes in GDP or the elasticity of tax to GDP should be set to be equal to or greater than one. If it could be achieved then this flexible tax could also assist in the anti-inflationary and anti-deflationary component of fiscal policy. So far the indicator of the Indonesian tax system’s responsiveness
to the change in income without adjusting the impact of discretionary changes (Indonesian tax buoyancy) shows relatively good figures compared to other developing countries. Using 11 years of panel data (1998-2008) for 25 developing countries, Ahmed & Mohammed (2010) calculated Indonesian total tax buoyancy of 1.5. It comprised of direct tax buoyancy of 1.78 and indirect tax buoyancy of 1.17. A tax buoyancy of one implies that a one percent increase in GDP tax revenue will match the increase of one percent, thus it will leave tax to GDP ratio unchanged. With figures greater than one, we could expect Indonesian tax revenue would increase more than its GDP ratio and potentially lead to reductions in the deficit ratio. The Indonesian tax system is among the middle countries compared with the other 25 developing countries. The buoyancy of 10 countries are higher than Indonesia, they are Bangladesh (2.58), Cuba (2.56), Ghana (2.37), Mexico (2.4), Egypt (2.19), and the rest were Cameroon, China, Iran, Sri Lanka and Uganda with total tax buoyancy between (1.59 – 1.89). However the elasticity was not calculated due to the unavailable data leading to difficulties to adjust the estimated impact of the discretionary tax changes for the 11 years (periods of observation).

5. Equity

Uppal (2003) reemphasized the important features of the efficient tax systems to improve the distribution of income in the society. However in the Indonesian condition of a very narrow tax base due to high rate of non-compliance and low tax-filing rate, an increasing rate is not a favourable option. In terms of tax equity, Heij (1993) strongly argued that without the chance to impose and enforce the laws equally to all eligible taxpayers, then the strategy to increase tax rates will contradict the goals of efficiency and equity of the tax. Clearly the problem of non-compliance and the very low ratio of tax return filing meant a recommendation to focus
more on administration improvement to be more efficient and competent (Asher, 1989; World Bank, 1992 and Booth, 1992). Furthermore, making this improvement could consistently increase the compliance, the administrative effort must be used with strong and persistent political determination. In addition to that, Gordon and Li (2009) underlined that the equity principle must be applied with equal treatment and application of the tax law.

Uppal (2003) emphasized that for more than three decades after its independence, the Indonesian government had never put any attention to restructuring the tax system and to use good tax policy to meet their major economic goals. Touwen (2008) subsequently reported the beginning of the Indonesian oil and gas-based economy. He stated that the exploitation of Indonesian oil had started to become profitable by the late nineteenth century. Since then, petroleum began to be one of the most important and also the main contributor in the total export packages of the Netherland-Indies (the former Indonesia under the Dutch Colony). Starting from its independence, the new Indonesian government continued with its dependency on the massive income stream from oil and gas exports. From the 1970s onward, the Indonesian new-order economy experienced significant windfall gains from the increased oil price on the world market. Figures from the government national budget show the higher dependency on oil and gas revenue in the report of the Indonesian five-years development plan III (Repelita III) covered the period of 1979/80 - 1983/84. The report from this medium-term development plan showed that 69% of domestic revenue for the period sourced from the oil and gas-related revenue (Heij, 1993).
Table 27. Comparison: Share of Oil and Gas Revenue and National Tax Revenue to Domestic Revenue

<table>
<thead>
<tr>
<th>Description</th>
<th>Average 1970's</th>
<th>Average 1980's</th>
<th>Average 1990's</th>
<th>Average 2005-2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of Oil &amp; Gas Revenue to Domestic Revenue (%)</td>
<td>47.20</td>
<td>54.35</td>
<td>27.05</td>
<td>27.60</td>
</tr>
<tr>
<td>Share of Tax Revenue to Domestic Revenue (%)</td>
<td>46.29</td>
<td>39.51</td>
<td>64.00</td>
<td>69.09</td>
</tr>
</tbody>
</table>

Sources: Ministry of Finance, 2009

However oil and gas resources depleted faster than initially estimated, especially from the 1970’s onward (as shown in table 27). This urged the Indonesian government to adopt a radical tax reform in 1983. On the other hand, the contribution of the overall tax revenue to the domestic revenue increased sequentially year by year. As explained by Anwar, Aziz, Pangestu & Soesastro (1991), Hill (2000) and Resosudarmo & Kuncoro (2006), the initial aim of the 1983 reform was to source alternative income streams for financing the declining trend of oil and gas revenue in the state budget. In addition to revenue adequacy as the main trigger, there was a strong push from technocratic finance and economic ministers behind President Soeharto that reforming the taxes will contribute to two key elements for Indonesian economic stability; economic growth and the elimination of poverty (Heij, 1993).

4.1.2 The detail of recent reform

The focus of this study will start from the 1983 income tax reform which effectively came into force in 1984. The main motivation of this reform was four fold. The first was to secure revenue from non-oil tax. Secondly to simplify both the income tax law and its administration in order to ensure more efficient and better services. Thirdly to improve income distribution by reducing tax-induced distortions in the allocation of resources and to achieve economic neutrality. The final aim of the reform was to ensure that a lower tax burden applied to the
poor (Asher, 1989, Gillis, 1985). The main features of the 1983 income tax reform was a broader tax base, significantly reduced tax rates and the omission of high income earners’ exemptions. Further, Gillis (1985) added that the applied rates within the 1983 law reform were considerably lower than in most less-developing countries (LDCs) and North America in these years. Table 28 details the history, in chronological order of the main changes featured in the Indonesian PIT reform since 1983.

<table>
<thead>
<tr>
<th>No</th>
<th>Income Tax Law</th>
<th>Annual Income Groups (in Indonesian Rupiah)</th>
<th>Tax Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No 7 Year 1983</td>
<td>Income ≤ 10,000,000</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10,000,000 ≤ Income ≤ 50,000,000</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Income &gt; 50,000,000</td>
<td>35%</td>
</tr>
<tr>
<td>2</td>
<td>No 10 Year 1994</td>
<td>Income ≤ 25,000,000</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25,000,000 ≤ Income ≤ 50,000,000</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Income &gt; 50,000,000</td>
<td>30%</td>
</tr>
<tr>
<td>3</td>
<td>No 17 Year 2000</td>
<td>Income ≤ 25,000,000</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25,000,000 ≤ Income ≤ 50,000,000</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50,000,000 ≤ Income ≤ 100,000,000</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100,000,000 ≤ Income ≤ 200,000,000</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Income &gt; 200,000,000</td>
<td>35%</td>
</tr>
<tr>
<td>4</td>
<td>No 36 Year 2008</td>
<td>Income ≤ 50,000,000</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50,000,000 ≤ Income ≤ 250,000,000</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>250,000,000 ≤ Income ≤ 500,000,000</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Income &gt; 500,000,000</td>
<td>30%</td>
</tr>
</tbody>
</table>

In Indonesia, the central government through the Ministry of Finance has the authority to determine both the tax bases and tax rates for the PIT. Nevertheless, amendments to law are necessary to make any changes in the income tax rate. These law amendments are a long process which require intense discussion between parliament and the government. In terms of PIT, much of the reform has been related to the income tax progression component. The Indonesian government first introduced a more progressive rate with a lower tax burden in
1994. This continued in 2000 when the government initiated an even more progressive tax rate. A big gap between the lowest and the highest tax brackets could create unfavorable incentives for income shifting or avoidance by the higher income earners (Yuwono, 2008). Those concerns were answered with the provision of the latest Law no 36(2008) which introduced more lenient tax brackets which makes its impact interesting for analysis in more detail in this study.

The less progressive and more lenient tax brackets were not the only things introduced by the new tax reform. Starting in 1983, the reform initiated an exemption for working wives as an incentive to encourage women’s participation in the labour force (Gillis, 1985). Following the first reform in 1983 and the second reform in 1994, more specific and comprehensive administrative reforms were initiated in 2000. This highlighted a starting milestone for tax administrative modernization (DGT, 2007). In the article 7(3), Law no 36(2008) allowed changes in personal exemptions to provide a tax relief. This is the authority of the Minister of Finance to propose the changes in the tax threshold to provide some relief for Indonesian’s that meet a certain condition to the People Representative Agency or the Indonesian parliament. This article aimed to add flexibility to the tax system to adjust the minimum cost of living and the development of the Indonesian economy in line with inflation. In the Indonesian PIT Law, the personal exemptions component is in the form of non-taxable income. In 2008, it had increased from previously IDR 13,200,000 per year to IDR 15,840,000 per year for the individual tax payer. This change also applied to the additional allowance for the tax payer’s marital status, working wives with joint tax returns, and dependent allowance (maximum of three dependents allowed for one individual tax payer). The detail of the history of changes from the first tax reform up until the 2008 is described in Table 29.
Table 29. Non-taxable income changes 1983-2008

<table>
<thead>
<tr>
<th>No</th>
<th>Law Base</th>
<th>Article</th>
<th>Individual Tax Payer</th>
<th>Married</th>
<th>Working Wife</th>
<th>Dependent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Law No 7 Year 1983</td>
<td>7 (1)</td>
<td>960,000</td>
<td>480,000</td>
<td>960,000</td>
<td>480,000</td>
</tr>
<tr>
<td>2</td>
<td>MoF Decree 928/KMK.04/1993</td>
<td>2</td>
<td>1,728,000</td>
<td>864,000</td>
<td>1,728,000</td>
<td>864,000</td>
</tr>
<tr>
<td>3</td>
<td>Law 10 Year 1994</td>
<td>7 (1)</td>
<td>1,728,000</td>
<td>864,000</td>
<td>1,728,000</td>
<td>864,000</td>
</tr>
<tr>
<td>4</td>
<td>Ministry of Finance Decree KMK 361/KMK.04/1998</td>
<td>2</td>
<td>2,880,000</td>
<td>1,440,000</td>
<td>2,880,000</td>
<td>1,440,000</td>
</tr>
<tr>
<td>5</td>
<td>Law No 17 Year 2000</td>
<td>7 (1)</td>
<td>2,880,000</td>
<td>1,440,000</td>
<td>2,880,000</td>
<td>1,440,000</td>
</tr>
<tr>
<td>6</td>
<td>MoF Decree 564/PMK.03/2004</td>
<td>1(1)</td>
<td>12,000,000</td>
<td>1,200,000</td>
<td>12,000,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>7</td>
<td>MoF Decree 137/PMK.03/2005</td>
<td>1(1)</td>
<td>13,200,000</td>
<td>1,200,000</td>
<td>13,200,000</td>
<td>1,200,000</td>
</tr>
<tr>
<td>8</td>
<td>Law No 36 Year 2008</td>
<td>7 (1)</td>
<td>15,840,000</td>
<td>1,320,000</td>
<td>15,840,000</td>
<td>1,320,000</td>
</tr>
</tbody>
</table>

By law, the Indonesian PIT is calculated on an annual base. The 2008 tax threshold of IDR 15,840,000 is 113.8 percent of the Indonesian annual average wages of IDR 13,900,000 (Indonesian Bureau of Statistics (BPS), 2008). Compared to other developing countries, it is among the highest tax relief provided. An OECD study by Gandullia, Iacobone & Thomas, (2012) compared PIT structures and burdens of several emerging strong economies, namely; Brazil, China, India, Indonesia and South Africa (BCIIS). They firstly explain that China provides a monthly allowance of 6.7 percent of the monthly average wages. Further, they explain that South Africa provides a basic tax credit of 84.2 percent from the average national wages to all taxpayers and an additional tax credit of 46.6 percent of the average wages to persons aged 65 and over. Then both Brazil and India applied a zero rated bracket in their progressive PIT schedule. For Brazil, this basic allowance is 105 percent of their average national wage, while for India the zero tax rate to gross income is 206 percent of their average national wages. In sum, India is the only comparative country that provides higher tax relief than Indonesia.

In addition to the tax allowances, all of those countries also provide additional family-based relief. Indonesia provides an annual family member relief of IDR 1,320,000. That is 9.5
percent of average national wages per individual member of the family. The maximum entitlement is for three dependents per tax payer with the position of a head of household. Others BCIIS applied various rates less than, or comparable with, that which is provided by the Indonesian PIT system.

4.1.3 The performance of PIT in Indonesia

Tax revenue contribution increased sharply during the two decades after the first major reform took place, from only 39.5 percent in the 1980’s to be at the average of 70 percent of the total government domestic revenue from 2005 to 2009. Despite this achievement, Indonesia still suffers from ineffective tax collection and low tax payer compliance (Ikhsan, Trialdi & Syahrial, 2005).

Although the ineffective tax collection and low compliance are common problems for the fiscal situation of developing countries (International Monetary Fund (IMF), 2011; Gandulia et al., 2012) the problem in Indonesia is considered bigger than neighbouring developing countries. Looking at how much tax is collected compared to GDP, figure 8 shows that Indonesia had the lowest total tax ratio to Gross Domestic Product (GDP) compared to the other Association of South East Asian Nations (ASEAN) countries. Of the ten ASEAN countries, the Indonesian tax ratio of 13.04 percent shared the fourth lowest position above Lao (12.10 percent) in 2008. Only Cambodia (10.56 percent) and Myanmar (5 percent) had lower tax ratios than Indonesia. Brunei (36.44 percent), Vietnam (26.43 percent), Thailand
(16.45 percent), Singapore (14.99 percent) and Malaysia (14.66 percent) were the countries with higher tax ratios in ASEAN countries.

Figure 8. Tax as a percentage of GDP; ASEAN Countries (2008)


Ironically, even with the low tax ratio, the contribution of income tax in Indonesia is smaller than other ASEAN countries with higher tax ratios. Figure 9 shows that income tax makes up at least 50 percent of the total tax collected in Malaysia and Thailand (Treasury Malaysia, 2010; Revenue Directorate (RD) Thailand, 2010). Although not as high as those two countries, Singapore and Vietnam still have around 40 percent of their tax revenue coming from income tax (Inland Revenue Authority Singapore (IRAS), 2010; GSO Vietnam, 2010). Therefore, the contribution of income tax at around 25 percent in Indonesia can be considered very small even without taking into account that the tax ratios of the other countries are higher than Indonesia’s tax ratio. Most of the income tax in these countries (an average of 33 percent of their total tax) comes from corporate income tax (CIT) while the PIT only
accounts for 15.40 percent of their domestic tax revenue (Treasury Malaysia, RD Thailand, IRAS Singapore, and GSO Vietnam, 2010).

![Figure 9. Tax mix comparison with neighbouring countries (2007)](image)

This contradicts the condition in developed and more advanced economies. In the case of the Organisation for Economic Cooperation and Development (OECD) countries, PIT share dominates with an average of 25 percent-32 percent from total tax revenue since the 1960s. Meanwhile, their average CIT share has been constant around 9 percent-10 percent.

Enforcement is the main problem in developing countries. The much lower contribution of PIT is also linked to the size of the informal economy in developing countries. The median size of the informal economy among the developing countries is 37 percent of GDP. It is much bigger than the estimated size of the informal sector in OECD countries which is 15 percent of the GDP (Gordon & Li, 2005).
In Indonesia, PIT contributed around 20 percent of the 41.2 percent total revenue that came from the non-oil and gas sector in 2009 (See figure 1 in chapter 1 that shows the share of tax revenue and detailed breakdown of income tax revenue in 2009). This means PIT only contributed around 8.2 percent of the total revenue while CIT has a much higher contribution of almost 20 percent of the total revenue (DGT, 2009). There are two types of PIT. The first is the withholding tax from employees’ with a single source of income, known as tax article 21. This tax accounted for 18.8 percent of the revenue from the non-oil and gas sector in 2009. The second one is applied to those with more than one source of income. It is known as tax article 25/29 and this contributed only 1.2 percent of total non-oil and gas income tax revenue. This study specifically looks at the PIT from labour-income (or PIT article 21) which contributes the major share of the Indonesian PIT.

Moreover, Indonesian minimum provision of non-taxable income is 7 times higher or 723 percent of the Indonesian per capita official annual poverty line and 4.5 times higher than the US$ 1 per day income poverty line suggested by the World Bank. It is more than twice or 2.3 times higher than a US$2 per day income poverty line. These arguments, combined with the Gandullia, et.al. (2012) findings, indicate that by looking at the their PIT structure, Indonesia, together with India and South Africa impose relatively low average and marginal PIT rates compared to the majority of OECD countries. Further, Gandullia, et.al. (2012) suspects that only a fraction of high income workers in those countries face a substantial tax burden and it could also explain their relatively low tax to GDP ratio compared to OECD countries. Our microsimulation model will compare, in detail, the different impact of different tax structures on income distribution, especially the one that based on the provision of Law no 36(2008).
4.2 Empirical Framework

To answer the main research questions, we developed PIT microsimulation for Indonesian individual tax payers. As previously elaborated in Chapter 3, microsimulation required the use of micro data to analyze the detailed distributional impact among the individual units in the datasets. It also enables the financial costing as an aggregate calculation of the policy impact (Mitton, Sutherland & Weeks, 2000).

As also outlined in the previous chapter, the majority of Indonesian income distribution studies used Susenas consumption expenditure as a proxy for the income distribution (Timmer, 2005; World Bank, 2006; Miranti, 2010; Suryadarma, 2010) while some have tried to apply limited use of the income information from the survey (Alatas & Bourguignon, 2005; Cameron, 2002; Leigh & van der Eng, 2009; Frankema & Marks, 2010). Further, Leigh & van der Eng (2009) showed that the less than adequate coverage of the Indonesian top incomes is the major problem, a serious limitation found from using the income base file provided by the BPS. Hence, it could be concluded that those previous studies only provide partial analysis of both Indonesian income and/or consumption distribution.

This study aims to further improve the reliability of the labour-income database by constructing the real income approximation to enable a full distributional analysis of the Indonesian income structure. As shown in chapter 2, we have combine the power of detailed characteristics and representativeness of survey data from both the National Socio Economic Survey and the National Labour Force Survey 2008 with high income coverage of tax administration data from the National PIT Return 2008 database. We apply microsimulation
to that database in order to assess the impact from the implementation of PIT Law no 36(2008). However, this study does not account for behavioral change. Therefore, the model is static and only measures the first-round effect just before individuals change their behaviour as a response to the changes. Behavioural changes can include changes like increasing or decreasing work hours or changing labour force status as a result of the policy change. This means we do not imply any standard assumptions to the labour supply elasticities in our labour income micro data set nor do we include any measure of tax evasion in our estimates.

After establishing the labour-income database, the tax estimator module was built using STATA. It generates output on revenue and the distributional impacts for every sampled tax payer in the database. We center our impact analysis on the taxpayers in different income deciles. The microsimulation analysis includes results for income tax liabilities, which represents government income tax collection from withholding labour income and the income tax burden across income deciles. Following Yuwono 2008, we calculate the tax burden as the income tax liability divided by annual gross income. We measured the difference in the tax burden before and after the implementation of the new law. We also measured the change in government income tax revenue, calculated as the new weighted income tax liabilities (under the proposed new law) minus the existing weighted income tax liabilities (under existing law).
4.3 Results and Discussions

4.3.1 The impact on the tax basis

This study aims to analyse the impact of the new tax law of 2008 on the tax capacity and distribution in Indonesia on one of the major features of the amendment - the change in tax structure. The new law has only four different tax rates as opposed to five rates of the income tax law 17/2000. As can be seen in the previous Tables 28 and 29, the new tax law has changed the non-taxable income together with income tax brackets and tax rates. Therefore, the microsimulation model estimates the potential tax revenue from these two tax structures.

Table 30 shows that the number of eligible tax payers has slightly decreased as a result of the reform. To take account of the income earners that are not eligible to be taxpayers, we used a non-taxable income group as one of the categories in Table 30 in addition to the formal income groups. This is to show the category of salary/wage earners with nil income tax liability. The application of the new law indicates a reduction of about 3.57 million tax payers or around 6.21 percent of total numbers of the potential personal tax payers. The increase in this category of non-taxable taxpayers is mainly a result of the increase in the income tax relief component with the application of the Law no 36(2008). This could be seen as a significant loss in the tax basis but given the people who no longer need to pay the tax are in the low income category and only need to pay the 5 percent rate, the actual value from the related tax revenue may not be substantial.
A larger impact of the law may well come from the considerable reduction of the tax rate for the top income groups. From the previous small fraction of 0.42 percent of income earners who paid up to the marginal rate of 35 percent in the highest income group, the new proportion reduces to only 0.11 percent who pay the new lower marginal tax rate of 30 percent and another 0.19 percent paying 25 percent. The remaining taxpayers paying a 35 percent tax rate (0.12 percent of the income earners) is now included in the much lower tax rate of 15 percent. Therefore, we should expect a significant reduction of tax revenue due to the decreasing tax basis and decreasing marginal tax rate.
4.3.2 The Impact on Revenue

Following the framework of analysis from Wallace, Wasylenko & Weiner (1991) for the US Tax Reform Act 1986 and Yuwono (2008) for the Indonesian Income Tax Reform 2008, we analysed the effects for both the changes in the taxable income base (the personal income tax relief) and changes in the tax rates. From both of these main changes from the personal income tax reform, we then assessed the impact on tax payers’ income tax liabilities and on tax payers’ income tax burden. Tax liabilities represent government income tax revenue collection from labour income. The change in tax liabilities is calculated as the new income tax liabilities (income tax liabilities under new income tax law) minus the current income tax liabilities (income tax liabilities under previous income tax law). The second impact on the income tax burden is calculated as income tax liabilities divided by gross annual income of the taxpayer. We simulate tax liabilities and tax burden under four scenarios: pre-reform (year 2000 income tax law) tax relief and tax rates scenario, pre-reform tax relief and post-reform (year 2008 income tax law) tax rates, the post-reform tax relief and pre-reform tax rates, and the post-reform tax relief and tax rates. In doing so, we can check whether most of the tax revenues and tax burden changes are attributable to the change in the tax bases or to the changes in the tax rates. Table 31 presents the results for the PIT simulation under our incidence assumption. Tax revenues are totalled while tax burden is averaged over the individuals in each tax payers’ income decile.
Table 31. Income Tax Revenue Simulation before and after Reform (Income Tax Denominated in Billion Rupiahs)

<table>
<thead>
<tr>
<th>Income Tax Decile</th>
<th>Before Reform</th>
<th>After Reform</th>
<th>Law 17/2000</th>
<th>Rates Change Only</th>
<th>Base Change Only</th>
<th>Law 36/2008</th>
<th>% from total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income Tax</td>
<td>% from total</td>
<td>Income Tax</td>
<td>% Change</td>
<td>Income Tax</td>
<td>% Change</td>
<td>Income Tax</td>
<td>% Change</td>
</tr>
<tr>
<td>1</td>
<td>92.90</td>
<td>92.90</td>
<td>0.00%</td>
<td>97.60</td>
<td>5.06%</td>
<td>97.60</td>
<td>5.06%</td>
</tr>
<tr>
<td>2</td>
<td>196.00</td>
<td>196.00</td>
<td>0.00%</td>
<td>179.00</td>
<td>-8.67%</td>
<td>179.00</td>
<td>-8.67%</td>
</tr>
<tr>
<td>3</td>
<td>280.00</td>
<td>280.00</td>
<td>0.00%</td>
<td>262.00</td>
<td>-6.43%</td>
<td>262.00</td>
<td>-6.43%</td>
</tr>
<tr>
<td>4</td>
<td>445.00</td>
<td>445.00</td>
<td>0.00%</td>
<td>406.00</td>
<td>-8.76%</td>
<td>406.00</td>
<td>-8.76%</td>
</tr>
<tr>
<td>5</td>
<td>659.00</td>
<td>659.00</td>
<td>0.00%</td>
<td>596.00</td>
<td>-9.56%</td>
<td>596.00</td>
<td>-9.56%</td>
</tr>
<tr>
<td>6</td>
<td>944.00</td>
<td>944.00</td>
<td>0.00%</td>
<td>796.00</td>
<td>-15.68%</td>
<td>795.00</td>
<td>-15.78%</td>
</tr>
<tr>
<td>7</td>
<td>1,270.00</td>
<td>1,270.00</td>
<td>0.00%</td>
<td>1,080.00</td>
<td>-14.96%</td>
<td>1,080.00</td>
<td>-14.96%</td>
</tr>
<tr>
<td>8</td>
<td>1,750.00</td>
<td>1,740.00</td>
<td>-0.07%</td>
<td>1,460.00</td>
<td>-16.57%</td>
<td>1,440.00</td>
<td>-17.71%</td>
</tr>
<tr>
<td>9</td>
<td>3,160.00</td>
<td>2,610.00</td>
<td>-17.14%</td>
<td>3,030.00</td>
<td>-4.11%</td>
<td>2,330.00</td>
<td>-26.27%</td>
</tr>
<tr>
<td>10</td>
<td>53,800.00</td>
<td>40,600.00</td>
<td>-24.54%</td>
<td>51,700.00</td>
<td>-3.90%</td>
<td>39,100.00</td>
<td>-27.32%</td>
</tr>
<tr>
<td>Top 5 %</td>
<td>49,800.00</td>
<td>37,600.00</td>
<td>-24.50%</td>
<td>47,600.00</td>
<td>-4.42%</td>
<td>35,800.00</td>
<td>-28.11%</td>
</tr>
<tr>
<td>Top 1 %</td>
<td>36,900.00</td>
<td>27,700.00</td>
<td>-23.90%</td>
<td>33,900.00</td>
<td>-6.87%</td>
<td>26,100.00</td>
<td>-28.30%</td>
</tr>
<tr>
<td>Total</td>
<td>62,596.90</td>
<td>48,836.90</td>
<td>-21.98%</td>
<td>59,606.60</td>
<td>-4.78%</td>
<td>46,285.60</td>
<td>-26.06%</td>
</tr>
</tbody>
</table>

Sources: own simulation

Table 31 shows the revenue impact of the tax reform 2008. It added further analysis to the previous section’s finding on the reduction of the tax base from the implementation of the new tax structure. It also shows that the reduction of high income earners who paid the higher tax rates is an even greater proportion. In addition to the changes in the tax base, there is also the second major component of the reform which had more of an impact; the changes to the tax rates. Because of the relationship between the components of the two changes, it can be expected that the potential tax revenue would be reduced by more than 6.21 percentage point of nominal reduction in the tax base (as shown in Table 31).

The result presented in Table 31 confirms the expectation. It shows that the potential tax revenue is reduced by a much higher rate than the reduction of the tax base itself. Our static microsimulation estimation shows that the total impact of the new reform is a 25.87 percent decrease in potential government revenue. To fully understand the total revenue impact, we further analysed the detailed breakdown of the 2008 reform. We assessed the impact of the
combined tax base changes and tax rates changes before finally concluding our analysis with the total impact from the 2008 reform.

As shown by the result for the first seven deciles, the change in tax rates generally did not reduce the revenue. There was no impact from the tax rate changes for those deciles. We can see that changes only started from decile 8 with a slight revenue reduction of 0.57 percent as a result of the tax rate changes. This indicates that the large number of low income earners from the first income brackets with 5% tax rates dominate deciles 1 to 7. In terms of the biggest percentage of revenue reduction from tax rate changes, the highest decile (decile 10) and the second highest decile (decile 9) experienced the biggest reduction. The tenth decile shows a 23.91 percent reduction, while the ninth decile experienced a 17.41 percent reduction in their potential tax revenue.

While the tax rate change produced no impact in the seven lower deciles, the tax base change produced a larger impact on the revenue reduction for these deciles. The large impact of the reduction started from the lowest decile 1 up until decile 8. We found that the change in the tax base impacted significantly and removed all potential revenue from income decile one and decile two. When we only change the base, almost all of the taxpayers in these deciles were actually not paying taxes. This is presumably due to a significant reduction associated with the omission of their taxable income base right after the application of the new tax exemptions based on the new legislation. The subsequent impact on the reduction of the income tax liabilities continued to be experienced in each decile. The higher the decile the lower the impact. We suspect the gradual revenue reduction from decile 3 to decile 8 could also relate to our previous findings of the impact due to the increasing tax exemptions in the
new legislation. This increased exemption resulted in a big shift for most of the taxpayers in the lower decile to be non-tax paying citizens.

The domination of the impact of tax rates started from decile nine and ten (including the top 5 percent and 1 percent) of the individual tax payers. Tax rates’ change reduced the revenue from decile 9 by more than 17.41 percent, while at the same time the change in the tax base reduced the revenue by about 13.9 percent. A more extreme impact came from the highest decile where the tax rates change contributes 23.91 percent while the change in the tax base results in not more than a 1.57 percent reduction. In sum, this last impact has the most extreme influence leading to a total reduction of 25.13 percent of revenue.

Yuwono (2008) found that under a full application of Law no 36(2008) (which was a scenario 4 in her study) tax payers in the lowest income deciles 1 and 2 and the highest income decile 10 contributed a larger share to the revenue compared to other income groups. Yuwono based her analysis on the administrative data of personal income tax returns of the previous year. Different from Yuwono’s findings, from the combined survey and administrative data, in our base file we found that under the application of the new tax law, all four highest deciles of taxpayers, starting from the lowest decile 7 to the highest decile 10, contributed a higher share of the potential tax revenue compared to the estimation of the previous law. Decile 6 contributed the same proportion of the total revenue. While the other 5 lower deciles (decile 5 down to decile 1) contributed a smaller potential tax revenue share after the reform. These findings were different to Yuwono’s findings which stated that although the total sum of income tax liabilities of the individual tax payers in each income decile were reduced, under the new law, the share of income tax liability for lower-income
decile increased. We found the opposing result, under the new law the share of income tax liabilities from lower income decreased while the share from higher income was increased. Together, those findings should indicate that income tax burden under the new law will be lower and more equally distributed.

4.3.3 The Impact on the Tax Burden Distribution

Table 32 shows the impact of tax reform 2008 on the distribution of the PIT burden disaggregated based on income decile of the tax payers.

<table>
<thead>
<tr>
<th>Income Decile</th>
<th>Before Reform Law 17/2000</th>
<th>After Reform Rates Change Only</th>
<th>After Reform Base Change Only</th>
<th>After Reform Law 36/2008</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tax Burden</td>
<td>Tax Burden</td>
<td>% Decrease</td>
<td>Tax Burden</td>
</tr>
<tr>
<td>1</td>
<td>0.0043</td>
<td>0.0043</td>
<td>-</td>
<td>0.0048</td>
</tr>
<tr>
<td>2</td>
<td>0.0107</td>
<td>0.0107</td>
<td>-</td>
<td>0.0029</td>
</tr>
<tr>
<td>3</td>
<td>0.0146</td>
<td>0.0146</td>
<td>-</td>
<td>0.0076</td>
</tr>
<tr>
<td>4</td>
<td>0.0180</td>
<td>0.0180</td>
<td>-</td>
<td>0.0117</td>
</tr>
<tr>
<td>5</td>
<td>0.0215</td>
<td>0.0215</td>
<td>-</td>
<td>0.0159</td>
</tr>
<tr>
<td>6</td>
<td>0.0244</td>
<td>0.0244</td>
<td>-</td>
<td>0.0194</td>
</tr>
<tr>
<td>7</td>
<td>0.0273</td>
<td>0.0273</td>
<td>-</td>
<td>0.0230</td>
</tr>
<tr>
<td>8</td>
<td>0.0308</td>
<td>0.0306</td>
<td>-0.03%</td>
<td>0.0269</td>
</tr>
<tr>
<td>9</td>
<td>0.0442</td>
<td>0.0350</td>
<td>-0.92%</td>
<td>0.0387</td>
</tr>
<tr>
<td>10</td>
<td>0.1261</td>
<td>0.0932</td>
<td>-3.29%</td>
<td>0.1217</td>
</tr>
<tr>
<td><strong>Top 5 %</strong></td>
<td><strong>0.1626</strong></td>
<td><strong>0.1194</strong></td>
<td><strong>-4.32%</strong></td>
<td><strong>0.1587</strong></td>
</tr>
<tr>
<td><strong>Top 1 %</strong></td>
<td><strong>0.2663</strong></td>
<td><strong>0.2033</strong></td>
<td><strong>-6.30%</strong></td>
<td><strong>0.2642</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.0662</strong></td>
<td><strong>0.0516</strong></td>
<td><strong>-1.46%</strong></td>
<td><strong>0.0609</strong></td>
</tr>
</tbody>
</table>

Sources: own simulation

Overall changes made to the PIT by Law 36 year 2008 had reduced the overall tax burden at the national level. These are in line with our previous findings about the decreasing revenue
impact as a result of the enactment of the new law. The highest individual income deciles (deciles 9 and 10) experienced a larger decrease in their total tax burden compared to other deciles. The burden reduction is even larger in the top five percent and top one percent of the individual income deciles. It seems that the reform largely eases the burden of the highest income decile while still maintaining a low burden on the lower income decile.

In total, tax rate changes contributed more to the reduction of the burden than tax base changes. However, as an impact from tax rate change, there was no change in the burden especially from the lowest income decile (one) to the seventh. The significant burden reduction only came from the minority of taxpayers in decile nine up to the highest decile 10. Apart from the result of little change due to the tax rate, we found different conditions for deciles one up to eight. For these deciles, tax burden changes were mainly from the base reduction, while for deciles nine and ten, the combined effects from the tax rate change and the tax base change meant a larger total burden reduction from the implementation of Law no 36(2008). We note that the tax base change does decrease the burden for decile nine and ten of the tax payers, however the burden reduction from the tax rates change was more dominant.

Summing up, implementing the new law will result in revenue reduction but will ease the burden for all income deciles. Moreover, the tax burden figure shows that it is the tax burden of the highest income earners that is reduced most by the new law. Nevertheless, the estimate still shows that the highest income decile still bears the highest burden. The highest decile bears 8.98 percent of the income tax burden, followed by 3.21 percent and 2.69 percent borne by decile 9 and 8, and so on. The lower the decile the lower they bear the income tax burden.
4.4 Tax potential versus tax compliance

So far, our microsimulation estimate shows a large decrease in the tax potential following the reduction of the number of tax payers and the rates they have to pay. However, there is another side to the story, which is the fact that less than 10 percent of Indonesian salary earners are registered by the tax authority (in 2008, total registered 5.43 million personal income tax payers from 57 million salary earners). Some taxpayers may not be registered and some tax may not be collectable due to the informal nature of some types of work but the potential for an increase in the number of taxpayers is there.

Ikhsan et al. (2005) argued that expansion of the tax base can be implemented given that the current tax base is very much concentrated in the highest income decile of tax payers. Our microsimulation estimate verified this finding; in 2008, the highest decile of tax payers paid around 86.69 percent of tax (Table 31). This exposes the potential to raise significant tax revenue through ‘extensification’ especially to the next lower income. Based on our microsimulation estimate, given the potential is there, the extensification program by improving compliance is more important at this moment than changing the tax structure to increase the tax basis or the tax potentials. Nevertheless, the estimate also shows that even with extensification the structure of income and tax will always preserve the concentration of income in the first and second highest decile. Thus, to some extent our microsimulation shows that the dependency on the higher income tax payers cannot be reduced unless there is a significant change in the distribution of income in Indonesia.
In the last part of this discussion, we argue that in 2009 the Indonesian government had considerable success in regards to the compliance in personal income tax payment. Our microsimulation has shown that the tax basis, and especially the potential tax revenue, has been decreased considerably as the result of the new law. However, the data shows that the revenue from personal income tax still increased in 2009. To answer this, we need to revisit the issue of under coverage of high income earners in the Indonesian income data. Hence we tried to explain the issue by constructing our 2008 income tax microsimulation base file based on the imputed Susenas data which added 2.2 million individual tax payers. Those tax payers’ data were extracted from the income tax article 21 (withholding income tax) return year 2008 which was submitted in 2009. The result from the addition of the 2.2 million imputed tax payers was to account for 70 percent of the estimated tax liabilities (around Rp 40 billion from the Rp 60 billion tax liabilities), calculated with our personal income tax estimator as shown in chapter III.

Table 33. Tax Ratio and Personal Tax Payers Compliance

<table>
<thead>
<tr>
<th>Details Tax Ratio</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil &amp; Gas Revenue to Domestic Revenue (%)</td>
<td>29.40</td>
<td>32.61</td>
<td>24.99</td>
<td>30.71</td>
<td>20.28</td>
</tr>
<tr>
<td>Tax to Domestic Revenue</td>
<td>70.26</td>
<td>64.32</td>
<td>69.53</td>
<td>67.26</td>
<td>74.06</td>
</tr>
<tr>
<td>Income Tax to Total Tax Revenue</td>
<td>28.43</td>
<td>26.05</td>
<td>27.54</td>
<td>25.58</td>
<td>30.90</td>
</tr>
<tr>
<td>Income Tax to GDP Ratio (% to GDP)</td>
<td>6.33</td>
<td>6.25</td>
<td>6.04</td>
<td>6.61</td>
<td>5.72</td>
</tr>
<tr>
<td>Tax to GDP Ratio (% to GDP)</td>
<td>12.51</td>
<td>12.25</td>
<td>12.43</td>
<td>13.30</td>
<td>11.61</td>
</tr>
</tbody>
</table>

**Personal Tax Payer Compliance**

| Registered Tax payers          | 2,564,735 | 2,959,006 | 3,251,753 | 5,431,689 | 8,807,666 |
| growth                         | 15.37%    | 9.89%     | 67.04%    | 62.15%    |
| Submitted Tax File Return      | 851,190   | 898,036   | 899,567   | 1,677,160 | 4,853,323 |
| growth                         | 5.50%     | 0.17%     | 86.44%    | 189.38%   |

Source: Own calculation using data from State Budget Realisation Report, Directorate General of Treasury, Ministry of Finance, the Republic of Indonesia, 2010 and Socio Economic Indicators, Bureau of Statistics (BPS) Indonesia, 2010
Table 33 shows the registered tax payers increased 62 percent from 2008 (5.4 million personal tax payers) to 2009 (8.8 million personal tax payers). The compliance rate for the submission of tax returns increased 189 percent which is more than three times the increase in the registration. Hence, if it is possible to construct the new base data for 2009, it should at least add another 3.4 million personal tax payers to the microsimulation base file. With the previous benchmark of more than 70 percent increased tax liabilities, the estimated potential liabilities by using the new base file will increase by at least 18.2 percent from estimated tax liabilities in 2008 (calculated from 70 percent increase of the estimated liabilities times 26 percent decrease from the application of the new law).

Table 34. Simulated Revenue using 2009 Compliance Base

<table>
<thead>
<tr>
<th>Income Decile</th>
<th>After Reform (simulated with 2008 Compliance Base)</th>
<th>After Reform (simulated with 2009 Compliance Base*)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Income Tax</td>
<td>% from total</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-0.26%</td>
</tr>
<tr>
<td>2</td>
<td>37.20</td>
<td>-0.08%</td>
</tr>
<tr>
<td>3</td>
<td>39.20</td>
<td>0.08%</td>
</tr>
<tr>
<td>4</td>
<td>203.00</td>
<td>0.41%</td>
</tr>
<tr>
<td>5</td>
<td>418.00</td>
<td>0.85%</td>
</tr>
<tr>
<td>6</td>
<td>703.00</td>
<td>1.43%</td>
</tr>
<tr>
<td>7</td>
<td>1,030.00</td>
<td>2.10%</td>
</tr>
<tr>
<td>8</td>
<td>1,500.00</td>
<td>3.06%</td>
</tr>
<tr>
<td>9</td>
<td>2,370.00</td>
<td>4.84%</td>
</tr>
<tr>
<td>10</td>
<td>42,900.00</td>
<td>87.56%</td>
</tr>
<tr>
<td>Top 5 %</td>
<td>40,100.00</td>
<td>81.84%</td>
</tr>
<tr>
<td>Top 1 %</td>
<td>30,500.00</td>
<td>62.25%</td>
</tr>
<tr>
<td>Total</td>
<td>48,997.00</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

PIT Actual Revenue 2009 52,107.34
% simulated revenue from the actual 136.11%

*) simulated with the assumption of 62.15% increase in compliance
Table 34 simulated the revenue impact from increase in 2009 compliance. We simulated the total revenue impact by using the 2008 tax payer’s base plus the addition of registered tax payers in 2009. The impact showed increased revenue in most deciles with the exception of decile one, which showed a slight decrease due to the addition of the negative revenue base. This negative revenue base is tax payers whose taxable income and calculated tax liabilities become negative after the subtractions of tax exemptions. The total impact is a 44.75% increase of potential revenue in 2009 or 36.11% higher from the realized/actual revenue in 2009. In addition, there is an increase in the share of income tax to total tax revenue from 2008 to 2009. Although the increase in the tax revenue from income cannot match the increase in GDP (this can be seen from the lower ratio of income tax to GDP), we argue that without increasing the compliance rate the ratio of income tax should drop much further to around 63 percent of the 6.6 percent of GDP or become around 4 percent of GDP. Table 33 supports this argument by showing that not only did the number of registered tax payer increase in 2009 this was followed by the number of those who submitted the tax return.

4.5 Conclusion

Our initial attempt to build a more accurate microsimulation model for Indonesian personal income tax has produced an analysis of the revenue and distribution impacts of the tax burden from the latest tax reforms. We conducted a simulation of the impact of main changes in the new law; marginal tax rate reduction, tax threshold increase, the simplification of four income groups from the previous five income groups and the increase of personal exemptions.
The result of our static microsimulation model shows that the new personal income tax policy implementation will yield significant losses to the potential tax basis and for the government potential tax revenue. However there is a favorable effect on tax burden distribution; all income deciles bear a smaller burden under the application of the new law. More equal distribution of the tax burden is achieved by the reduction of the income tax burden in the highest income decile while still maintaining a low burden in the lower income decile.

Based on the findings of a high concentration of income and tax in the highest decile, our microsimulation estimate verifies that the structure of income and tax will always preserve the concentration of income in the first and second highest decile. This, to some extent, shows that the dependency on the higher income tax payers cannot be reduced unless there is a significant change in the distribution of income in Indonesia. Because of this, we believe that the ‘extensification’ program by improving compliance is more important at this moment than changing the tax structure to increase the tax basis or the tax potentials. The main features of the new law provide a favorable incentive to increase tax payer voluntary compliance. The revenue and compliance figure in 2009 shows that the government had considerable success in their first step to increase the tax basis and personal tax payer compliance.
CHAPTER V. Conclusions and Findings

5.1 Introduction

The initial motivation of this thesis was the need for an alternative model to provide detailed impact for income tax policy analysis. This need started from what has been observed in the Indonesian tax policy process over the years. In Indonesia, typically income tax law amendment is a long process involving intensive discussion between the Indonesian parliament and the government. The last 2008 amendment took almost four years to finalise. Therefore it is understandable why there have been only three amendments to the personal income tax law since 1983, despite the performance and changing condition of the Indonesian income tax and the economy.

Macro analysis plays a major role in supporting government tax policy and law amendments’s discussion. The main macro analysis emphasizes the interaction between macro factors and government policy intervention, and also the impact of global and domestic economic performance. This macro aggregate level analysis is cost effective and suitable when the primary area of concern is the tax impact of tax policy on the macro-aggregate level.

However, in order to satisfy the growing public awareness about the importance of taxation in the Indonesian current democracy, microsimulation models are the best policy tools. These
models can answer the need for detailed information on the impact of several policy scenarios (Citro & Hanushek, 1991). Microsimulation can provide an analysis of the interaction between individual components of the proposed system and an assessment of the relational impact of the proposed policy with other welfare programs. It is a more appropriate model for supporting evidence-based policy to satisfy the needs for transparent and accountable tax policy reform.

Since the initial Indonesian tax reform in 1983, the total tax ratio to GDP and the compliance rate indicate minor improvements despite the increasing value of revenue collection. This shows that the Indonesian tax collection, tax administration and compliance program still needs some refinement. Ongoing reforms need to be aimed at building an efficient tax administration and tax enforcement system without neglecting public service obligations. This refinement will create winners and losers. The winners are individuals or firms who experience less of a tax burden after reform, and vice versa. The PIT microsimulation model could also be used to contest the connection between high income contributions to tax revenue and the Indonesian income distribution. This study estimates the potential revenue impact, analyses the winners and losers from the policy, and examines the impact of the reform on the distribution of the tax burden.

This microsimulation approach is the third attempt to analyse Indonesian tax performance. Previous Indonesian studies include Marks (2003) who analysed the potential PIT using 2002 data and Susenas and Yuwono (2008) who built a PIT microsimulation model using administrative tax return data as the base file. This study will improve the Indonesian personal income tax base file for microsimulation input. It will contribute a static PIT which
will be more reliable and answer the main research questions of the thesis. In addition, the microsimulation model built can be used by the Indonesian tax authority for analysing the impact of policy reform.

This chapter summarises the findings from this thesis, and then outlines some limitations of the final model.

### 5.2 Findings

#### 5.2.1 Building Indonesia income base file

The strength of Susenas is in the richness of its information. The demographic characteristics and socio-economic variables of the sampled household, has made Susenas the main data source for most income and/or consumption-based studies. The richness of household and individual characteristics in the core survey, combined with the module data on consumption, expenditure, income, health, and education have made Susenas a rich data source.

Some scholars note a limitation on the coverage of poorest household as well as Susenas’ ongoing inability to cover the richest decile of the Indonesian households (Cameron, 2002; Leigh & van der Eng, 2009; Yusuf, 2011). This means Susenas still needs to be improved to allow analysis of income/non-income poverty, income distribution, taxation and other income based-studies.

Susenas also suffers from under coverage of high income respondents, under reporting of income, and under enumeration of income data. The main cause of this is the voluntary
nature of respondent participation and often difficult access to Indonesia’s luxurious housing complexes (BPS, 2010). In addition to that, Groves and Couper (1998) observed that high-income households might be less willing to be involved in such surveys because of their privacy, time and opportunity cost considerations. These limitations are generally in line with findings from studies of other advanced countries. Several studies of micro data quality, and assessment from the existing microsimulation modelling experiences for advanced countries like US (Radner, 1981; Citro & Hanushek, 1991), Germany (Wagenhals, 2004), UK (Eason, 1996), Canada (Naylor, 2000) also note their concerns about limitations of the income survey data.

Previous studies (Suryahadi, Sumarto & Maxwell, 2001; Filmer & Lindauer, 2001) show comparable income data and labour characteristics in Sakernas, our second data source. Thus we believe Sakernas labour income variables will be the best source of incomes to improve the Indonesian income base data. Citro & Hanushek (1991) emphasized the role of overlapping survey data to provide a benchmark for data validation. Therefore, variables like demography, socio economic status and, in particular, labour income in Sakernas will be the best variables to validate the microsimulation base file. We will also exploit the Sakernas labour income data to study the personal income tax and income distribution.

The third and last source for our base file is the income tax administrative data. It has the advantage of detailed income information from all taxpayers who submit their tax return and more readily available information for tax calculation based on the Indonesian income tax law. The nature of the administrative data, which is richer in income information, is a key addition to the survey data (Citro & Hanushek, 1991).
However the administrative tax return database has less coverage of low income earners and does not cover non-taxpayers. Currently there are several limitations to using Indonesian tax administrative data - the data are not weighted and there is no pooled cross-section dataset for all registered taxpayers for the observed period. This is due to a low compliance rate on tax return submission. The Indonesian tax payers’ compliance only averaged between 35% and 40% before 2008 (DGT, 2010). We also observe some missing values and variables in our tax return data due to manual recording errors. These make it difficult to conduct an accurate analysis of the Indonesian potential tax revenue. Hence, we choose to maximize the usefulness of this administrative data by using it to provide information on the top income individuals.

We used the framework for data matching from D’Orazio, Di Zio & Scanu (2006), which was modified in Eurostat (2013). The first step for data matching is the harmonisation and reconciliation of the data sources. They highlight eight stages for the reconciliation process: harmonisation of unit definition, reference period, target population, variables and classifications and then adjustment for accuracy, missing data and derivation of the variables. We modify the process by first reconciling our target population. We then proceed to perform a comparison of concepts and investigate the distribution of the common variables.

To assess this new dataset, we compared the plots of Sakernas, Susenas imputed, and the personal income tax return database. The plotting results confirm the need to add the higher income earners tax returns. The findings complement the previous concern about the lack of coverage of high income earners. It has long been considered an inherent weakness of the available Indonesian income micro data. Our validation results confirm a significant improvement to the coverage of the higher income earners. However, we have another
pending problem in the lower number of income earner observations from our Susenas imputation. This lower number of income earners will lead to a lower number of eligible taxpayers in the imputation base file.

### 5.2.2 Building Indonesian PIT microsimulation

The tax microsimulation has been built with the various datasets that we have. By only applying the tax simulator to the unmodified survey micro data we can extract information on the characteristics of taxpayers disaggregated by decile of income. However, the application of the tax simulator to the original database shows how the data base is lacking in observations of high income earners and cannot reach the potential total tax liabilities that can be achieved by using the imputed databases. We found that almost 71 percent of the income tax is contributed by the highest income earners in decile ten. In addition to this high concentration, the end result of the total tax liabilities indicated a significant underestimation of the income tax from labour/employment. This also indicated a problem of a low number of observations of high income earners from the Sakernas database.

The major benefit of having a static tax simulator is its ability to assess the distributional impact of the gains or losses for each micro-unit from future policy. There are several variations of the rules of PIT eligibility that can be simulated. The impacts can also be assessed within specific small units of analysis (for example, within certain thresholds of low income earners in particular municipalities or looking at the tax payer with two income earners and three children in the household) or further aggregated units of analysis (e.g., the national population).
Another benefit is that we can use our tax estimator to assess how different factors will impact on different deciles of tax payers. Our assessment of how the application of tax exemptions and the number of dependents impact on the reduction of the income tax liabilities showed that the result from Susenas-Sakernas-PIT imputation data has a similar pattern to the results from Susenas-Sakernas imputation.

The tax microsimulation model was then used to analyse several tax reform scenarios. We found that by applying different maximum dependent entitlements, the government will consistently yield more revenue if the stipulated tax law allows for fewer dependents in the tax relief component. We also found a supporting impact on the revenue; the stipulation of lower dependent deductions will increase the burden on the tax payers in general.

Another policy simulation assessed the impact from the application of increases in the taxable income threshold. The simulation started with a 1 million rupiah increase in each of the taxable income thresholds, then a 2 million rupiah increase, a 5 million rupiah increase and, lastly, an increase of 10 million rupiah. Interestingly, the impact on the revenue yield and the average effective tax rates are similar for most deciles (decile one up to decile eight). It appears that the real impact starts to happen for tax payers in decile nine and ten. The application of higher taxable income thresholds seems to decrease the government revenue and, in line with the decreasing the revenue yield, the tax burden of the taxpayers is also decreased.
5.2.3 Revenue and distribution impact (PIT reform in 2008)

In an empirical chapter we apply the static microsimulation model to real reform that took place in 2008 and was implemented in 2009. The result of our static microsimulation shows that the new personal income tax policy implementation will yield significant losses to the potential tax basis and government potential tax revenue. However there is a favorable effect on the tax burden distribution; all income deciles bear a smaller burden under the application of the new law. A more equal distribution of tax is achieved by the reduction in the income tax burden in the highest income decile while still maintaining a low burden in the lower income decile.

Our results from a microsimulation model verify that the structure of income and tax will always preserve the concentration of income in the first and second highest decile. This, to some extent, shows that the dependency on the higher income tax payers cannot be reduced unless there is a significant change in the distribution of income in Indonesia. Because of this, we believe that the ‘extensification’ program, by improving compliance, is more important at this moment than changing the tax structure to increase the tax basis or the tax potential.

Revenue and compliance figures in 2009 showed that the government had considerable success in their first step to increase tax and personal tax payer compliance. In the near future we will improve our model to simulate detail compliance of each person in each income
decile and to examine the potential for the government to increase the tax basis, the compliance rate, and the income tax revenue generation.

5.3 Implications for Policy

In the end, there are lessons from this study that can have major implications for Indonesian tax policy. The first lesson is that the study has shown how important tax data are to build the microsimulation model. Therefore, it is important for the Ministry of Finance to develop the administration, recording and reporting systems for tax data. This dataset should be updatable and can be used within the ministry for the analysis of the alternative tax policies.

Building the database also highlighted an issue about the income estimates from the national survey, especially regarding estimates of incomes for high income earners. This study shows that the current data could only capture at most 30 percent of the potential tax. As a consequence, it is important for BPS and tax authority to share data and try to fill the gap in Indonesian income data. This will be important not only for the study of tax estimates but also for other studies regarding inequality and poverty that also have implications for the welfare policy of the central and local government.

Despite major issues in the databases, this study shows that tax microsimulation for Indonesia can be built. This study shows the potential importance of this tool for the Ministry of Finance especially in regards to a reasonable target for tax maximisation. The microsimulation shows that the tax office still has to rely on earners in the highest two or three income deciles for tax revenue. This confirms the importance of information about
people in these deciles. Pursuing tax for lower income earners will less likely achieve a substantial increase in tax revenue.

The main aspect that can lead to substantially higher tax revenue is increasing compliance of the higher income earners. The success of the 2008 tax reform that increased tax while dropping the tax rate of high income earners confirms the importance of this strategy and compliance is currently more important than increasing the tax rate.

5.4 Limitations of the Current Study

5.4.1 Problems of the Indonesian Income Data

One of the major issues in the estimation of the lower and higher incomes in Indonesia is the unavailability of good income data. This has been an ongoing problem in the study of Indonesian income distribution (Cameron, 2002; Leigh & van der Eng, 2009). As mentioned in the previous chapter, most of the studies on poverty and income distribution used consumption expenditure data to proxy the income in Indonesia. The availability of good income micro-data is not only a crucial starting point to better understand the structure of Indonesian income but also to understand the impact of government policy intervention, such as their tax-benefit policy reform, on income redistribution.

We attempt to maximise the contribution of each data source by building a Personal Labour-Income Tax base file from a combination of the available Indonesian income survey and administrative micro data. However we also accept that each data source has its own problems and limitations due to the nature of data collection. Our preliminary analysis
indicates that the Indonesian survey data suffers a lot from under-reporting of income and under-reporting of high income respondents. Based on discussions with an officer from the Indonesian Bureau of Statistics (BPS), this is caused by the voluntary nature of respondent participation and by the limited access to several luxurious housing complexes. Indonesian administrative data particularly suffers from being non-weighted data. It also suffers from incomplete coverage of taxpayers due to low submission of taxpayers’ returns. We also found some missing values and variables due to manual recording errors. There is therefore a challenge for us to improve the income data base for our income microsimulation model by using other available data sources.

Based on our result in chapter II and III, in order to solve the data problems, we are required to adjust the data by using various techniques of data matching. Our study combined each of the available micro data using statistical matching and imputation. The main base data for was derived from the Indonesian Social Economic Household Survey, Susenas 2008. This Susenas was combined with Sakernas 2008 for labour-income completeness and for the weighting process. DGT administrative data of reported personal income tax returns from 2008 augmented the higher income coverage thus enhancing the survey data.

Using several matching variables, the missing Susenas labour income data was imputed from Sakernas. The new Susenas-Sakernas base file was then augmented with the high-income groups from the income tax administrative data. When this linked income base file was finished, non-taxable income estimation was conducted. A gross-up equation was built to identify individuals who earn gross salaries above a taxable income threshold under both the previous law no 17(2000) and the latest Law no 36(2008). This is the final base file which
produced an original parameter dataset, which was based on Law no 17(2000) and a changed parameter dataset, which was based Law no 36(2008).

5.4.2 The issue of compliance

The main limitation of this research is the issue of compliance. There are many questions that still remain un-answered surrounding compliance. As has been elaborated by Ahmed and Donoghue (2007), the accuracy level of microsimulation models would be greatly improved if the consideration of tax evasion and non-take up of benefits could be estimated. As stated in chapter IV, based on our microsimulation estimate, the extensification program by improving compliance is more important at this moment than changing the tax structure to increase the tax basis or the tax potentials. However, as far as we could observe, there are still few studies (including by the DGT administration) regarding the level of non-compliance. The administrative compliance rate published by the DGT was only a measure of the number of tax filers, without taking account of the tax payer who pay taxes but did not file their return. The availability of this information would be useful to help evaluate our argument about the compliance issue. The current administrative measure of compliance is not an estimate of the possible increase in tax payers and their level of income. This is something that is viewed as necessary for the future.

5.4.3 Static microsimulation to answer the research questions in this study

In order to assess the impact of the implementation of PIT Law no 36(2008), we build a static microsimulation model. Static microsimulation models are most often used to examine the potential impacts of detailed changes to tax and transfer policies. Static microsimulation
models are used to simulate rules based systems by replicating the set of rules to certain attributes of the individual micro-units. As elaborated in this thesis, static microsimulation recalculate the eligibility of micro-units to pay taxes based on the set of stipulated rules/laws.

Our static microsimulation model only measures the first-round effect, just before individuals change their behaviour as a response to the policy change. We do not imply any standard assumptions about the labour supply elasticity in our labour income micro data set nor do we include any measure of tax evasion in our estimates.

As Harding (1996) explained, ageing techniques are critical features that differentiate static and dynamic models. Static microsimulation models use two basic steps for data ageing; reweighting and uprating. Reweighting, sometimes also called ‘grossing up’, involves adjusting the weight of each individual in the household in the dataset. Reweighting usually ages the sample to bring it up to date or up to less than three years in the future whereas uprating relates to monetary value adjustments to reflect fluctuations from the original date of the survey. However, some models like ours might not need ageing as they are being used for analysis within the same period that the data were collected. Hence uprating is what we apply in all 2008 survey and administrative data to assess the impact of the 2008 tax reform. After uprating, a set of computational programs adopting the same rules as the government tax policy are applied to each of the individuals in the dataset. The result of the set of present rules is compared to the set of future scenarios. The end result of our static microsimulation model is a measure of the immediate impact of the policy change or the “morning after impact” with the assumption that there is no behavioural response from each of the individuals in the observation.
References


*Survei Sosial Ekonomi Nasional (SUSENAS) 2008 (micro data).* Available from BPS Indonesia.


*Trends of the Selected Socio-Economic Indicators of Indonesia 2009.* Indonesia: Author.


*Trends of the Selected Socio-Economic Indicators of Indonesia 2010.* Indonesia: Author.


*Trends of the Selected Socio-Economic Indicators of Indonesia 2011.* Indonesia: Author.


*Trends of the Selected Socio-Economic Indicators of Indonesia 2012.* Indonesia: Author.


Cameron, L. (2002). Growth with or without equity? The distributional impact of Indonesian
development. *Asian Pacific Economic Literature*, 6(2), 1-17.


dikenakan pph final dan/atau bersifat final, 4. dari penghasilan lain (formulir 1770).


Eurostat (2013). Statistical matching of EU-SILC and the Household Budget Survey to
compare poverty estimates using income, expenditures and material deprivation.


Haig, Robert M. (1921). The concept of income—economic and legal aspects (pp. 1–28). *The


Heij, G. (1993). Tax administration and compliance in Indonesia, Western Australia: Asia Research Centre on Social, Politics and Economic Change, Murdoch University.


