Methodology for Creating the RIGA-L Database

Esteban J. Quiñones, Ana Paula de la O-Campos, Claudia Rodríguez-Alas,

Thomas Hertz and Paul Winters[1]

Prepared for the Rural Income Generating Activities (RIGA) Project[2]

of the Agricultural Development Economics Division,

Food and Agriculture Organization

December, 2009

This document explains the methodology utilized in creating the RIGA Labour Database (RIGA-L). Details on issues of specific countries in the database are included in appendix II. For more information about the RIGA project, please refer to http://www.fao.org/es/esa/riga.

---

1. Introduction

As part of a broader project to examine the income generating activities of rural households across a range of developing countries[3], FAO has embarked on a study focusing on the wage employment activities of rural individuals. The broader project—referred to as the RIGA (Rural Income Generating Activities) project—among other activities has created household-level income aggregates using a consistent methodology and surveys from more than 15 countries. Along similar lines, the wage employment component of the RIGA project that is discussed in this paper seeks to create data on the labor market activities of rural individuals.

As in the component to create household income aggregates, a critical element of creating rural labor market data includes identifying comparable variables for analyzing labor market activities. Two areas of particular importance to consider is how to categorize the time spent working in

---

[1] Esteban Quinones is an Economist at the International Food Policy Research Institute, Washington D.C.; Ana Paula de la O-Campos is an Economist at the Food and Agriculture Organization of the United Nations, Rome, Italy; Claudia Rodriguez-Alas is D.C. Policy Office Director at SHARE Foundation, Washington D.C.; Thomas Hertz is a Consultant for FAO-ESA; and Paul Winters an Associate Professor both at American University, Washington, DC.

[2] The RIGA Project is a collaboration between FAO, the World Bank and American University in Washington, D.C. Original data can be obtained from the World Bank's Living Standards Measurement Study by visiting the LSMS website at: http://www.worldbank.org/lsms.

[3] The broader project is referred to as the RIGA (Rural Income Generating Activities) project and information on the project can be found at http://www.fao.org/es/ESA/riga/.

labor activities and how to determine earnings (wages) from those activities so that comparisons can be made across industries for individuals within countries as well as across countries. This is complicated by the fact that labor market modules differ across surveys in the manner in which they collect information and often cover different time periods within a given year.

The purpose of this document is to explain the methods used to create the RIGA Labor (RIGA-L) data for the country surveys included in the data base. In addition, it is intended to provide a guide for researchers on how to use the RIGA-L data. This is accomplished first by addressing the issue of labor time use. Once this is clarified, the manner in which labor earnings are considered is more straightforward. As such, Section 2 details the methodology for classifying jobs according to labor time and Section 3 presents the approach for estimating monthly earned income and daily wages. Section 4 discusses individual-level characteristics which were created to analyze labor market participation and wage earnings. Finally, Section 5 details how to use the RIGA-L data. While this provides the overall method, further information on individual countries as well as the specifics of using the data set are provided. In particular, Appendix I provides basic information about each survey, as well as a list of variables that are generated for each country, while Appendix II details country-specific issues that came up while creating the data base and the actions taken to deal with these issues. Finally, an employment income aggregate and individual characteristics technical note is attached in Appendix III which provides details on the organization and use of the data.

2. Creating Labor Time Variables

The prevailing labor time characteristics of jobs in the labor market are of particular interest since they indicate the degree to which an individual is involved in the labor market and because they are likely to influence participants' earnings.[4] Key areas to consider in creating time variables include: A) Do individuals engage in full time labor or are they employed in one or more part time jobs? B) Are laborers engaged in year round work, seasonal labor, or intermittently available casual work? Categorizing types of employment can be a thorny and complex process and the best approach often depends on the data available from a particular data set. However, since our interest in creating the RIGA-L database is in cross-country comparisons, it is important to take an approach that is applicable across a range of situations. As a result, a simple method focused on answering the questions above is relied on to create a clarifying and manageable framework.

In particular, all employment is categorized into one of the following four classifications: A) Full Year-Full Time (FYFT), B) Full Year-Part Time (FYPT), C) Part Year-Full Time (PYFT), and D) Part Year-Part Time (PYPT). These groups are intended to capture the labor time characteristics of individual employment and reflect the predominant types of jobs that exist. It can be assumed that the FYFT category represents full-time employment while the FYPT category represents part-time jobs. In addition, the PYFT category represents seasonal jobs and the PYPT category represents causal employment. Of course, there is a final category of individuals; namely those that do not participate in the labor market.

[4] Also of interest are the industries and occupation categories that labor opportunities fall into as they too are likely to define returns to participants. This topic is treated in more detail in section 3.2.

The next part of this document provides an explanation of the determinants for each group and the assumptions they are based on. Following this discussion, the particularities and challenges encountered in this process are detailed.

2.1 General Principles

There are two labor time dimensions which define specific employment that are of particular interest: duration and frequency. *Duration* is the length of time that a job has continuously been worked at by a specific person in a given time span, such as the number of months worked in the last year. The duration of a job can be considered as short as one day to as long as one year. *Frequency*, on the other hand, refers to how often a job is worked at by an individual in a given time span, such as the number of hours per week during the duration of a particular job. Frequency can include a few hours a day or a few days a week up to full hours a day and a full week's work. The duration of a job is an important issue to consider because it provides an understanding about the stability of the employment, as well as the continued opportunity it provides the employee to earn income over time. Both the duration of a job and the frequency of work may also influence the level of wage compensation provided in return for supplying labor. The frequency of work is essential to consider because it is likely to affect an employee's ability to work in other jobs to earn additional income. In combination, duration and frequency, along with details concerning the type of job and industry, play a considerable role in defining earnings.

Labor time is commonly specified in the following units: years, months, weeks, days, hours, etc., and combinations of those, such as hours per month. As such, the duration and frequency of work can be defined in a number of ways, depending on the units in which this information is reported. Duration is best defined over the year since there is a tendency to think of the timing of work over a year (e.g. seasonality implies changes over the year). Since the next smallest time unit is months, months per year makes an ideal measure of duration and provides a sense of the longevity of work over a given year. In terms of frequency, to get sufficient detailed information it is desirable to use relatively short time periods for a base, such as the work week, in order to get a better idea concerning repetitiveness. Ideally, frequency would be taken into account as hours per week.

However, it is common to find different time units in labor modules of different surveys, which makes it challenging to use a standard set of time units. Consequently, it is not always possible to take into account months per year and hours per week for each survey and alternatives must be relied on. In addition, even within a single survey labor time questions are not always consistent. For example, in a specific survey complete labor time information may be available for main and secondary jobs, but incomplete for third or casual jobs. Again, in these cases alternative measures must be implemented.

2.2 Applying the Framework

Given the aforementioned concepts and practical complications, a standard framework is provided. This methodology is essentially a set of rules for classifying employment according to

labor time information; from time to time, exceptions to these rules do occur because of country specific situations or a lack of sufficient labor time information in a survey. It should be noted that the framework has been formulated not only to take into account the information above, but also to minimize the number assumptions necessary to proceed. Therefore, the A) FYFT, B) FYPT, C) PYFT, and D) PYPT classifications are based on the following assumptions:

Duration:
- *Full Year job* >= 10 working months per year
- *Part Year job* < 10 working months per year

Frequency:
- *Full Time job* >= 35 hours per week
- *Part Time job* < 35 hours per week

In combination, labor participants are grouped into one of the following four categories depicted in the table below[5]:

Table 1. Labor Time Matrix

	Duration	
Frequency	*FYFT:* >=10 months & >=35 hours	*PYFT:* <10 months & >=35 hours
	FYPT: >=10 months & <35 hours	*PYPT:* <10 months & <35 hours

Since surveys vary in their labor reporting of the timing of information often does not exactly match the time categories defined above. To make them comparable, in the absence of "months per year" and/or "hours per week" questions, the following methods were used to determine duration and frequency classifications.

Table 2. Methods to Determine Duration and Frequency Classifications

	Methods
Duration	- Months: If the number of months is not available, the number of days per year is divided by days per month to estimate the number of months per year worked.
	- Weeks: In the absence of months per year, weeks per year are used to designate full year or part year employment. It is estimated that 44 weeks are equivalent to 10 months. - Another way to determine the number of weeks worked per year is multiplying weeks per month by the number of months worked.
Frequency	- Hours per week: If the number of hours per week is unavailable, we

[5] These categories were created in the absence of a pre-existing set of guidelines. An established methodology for categorizing types of employment was searched for in a variety of resources, such as the website of the International Labor Organization, but none was identified.

divide hours per month by 4.35 (the estimated number of weeks per month) to get hours per week.

Days per week: In the absence of hours per week, days per week are used. Five days or more per week are assumed to designate full time status and less than five days per week as part time status.

- When days per week are not available but hours per day are available, hours per day are multiplied by the number of days worked in a week.[6]

2.3 Data Issues

In applying the framework, a variety of specific data issues are encountered. Although only affecting a small fraction of total observations, it is necessary to deal with these issues to avoid the loss of observations in the data and to create consistency in the data sets. The main issues encountered are discussed in this section.

2.3.1 Missing Values & Outliers

Before categorizing observations as described above, the labor time variables are checked for missing values and outliers. When missing time values exist, and it appears as though they should not be missing (based on the values of other time, wage participation, and income variables), they are replaced with the median of non-missing and non-outlier observations. This procedure rarely affects more than a handful of observations for each survey and is preferable to leaving the values as missing for two reasons: 1) leaving the value as missing will exclude these observations from our categorization and may exclude the observations in future analyses, and 2) leaving the value as missing may falsely assume a value of zero once the data is collapsed to the appropriate level.

In the case of time variables, an outlier is defined as an observation with a value outside the range of possibility, i.e. 13 months per year or 8 days per week. These instances exclusively refer to values that are too high, not those below a certain range. In these instances, values are recoded with the maximum possible value, instead of the median. Although the existing values are erroneous, it is more appropriate to replace them with the maximum than the median, because it is assumed that the true value of these observations is at or closer to the maximum possible value than the median of the distribution. Below is a list of maximums used, followed by a brief explanation when warranted:

- *Months per year*: 12 (the maximum per year).
- *Weeks per year:* 52 (the maximum per year).
- *Weeks per month*: 4.35 (365.25[7] days per year divided by 12 months, all of which is divided by 7 days per week – ((365/12)/7) – which rounds to 4.35).

[6] In general, the median days per week worked for job 1 is six in most countries/surveys (the means and modes also hover directly around six). As a result, when days per week are not available six days per week is relied on for the purpose of facilitating analysis.

[7] 365.25 is used in calculations, instead of 365, to account for the extra day in the calendar every 4 years (leap year)

- *Days per year:* 365 days per year (the maximum per year) or 312 working days per year (52 weeks multiplied by 6 working days per week. This is used if it is more appropriate for a select survey).
- *Days per month*: 31 (the maximum per longest month).
- *Days per week*: 7 (the maximum per week or 6 working days per week, if more appropriate for a specific survey).
- *Hours per day:* 16 (assuming that an individual can work a maximum of 16 hours in a single day).[8]
- *Hours per week*: 84 (assuming that an individual can work a maximum of 12 hours per day for 7 days or 14 hours per days for 6 days, etc.). Note: this implies that it is not possible for an individual to work for the maximum number of hours per day, 16, for more than 6 days.

2.3.2 Job Discrepancies

One of the inherent challenges in a multi-country study is the differing ways that individual surveys ask labor time questions. For some countries, labor time questions differ according to first, second, or third jobs; while in some cases all of the labor time queries are consistent. In addition, in some countries the first job is designated as the primary or full-time job whereas the second job is considered as casual, other, or default employment. This can be problematic when a person has two full-time jobs, or when a person has no full-time job but two part-time jobs or more. In such a case, it can be difficult to designate one particular employment as the primary or secondary job; varying criteria can be applied to decipher this, such as labor time or earnings, which can be further complicated when labor time or income questions are not consistent throughout employment modules.[9] Another aspect to consider is that some surveys request information for only main and secondary jobs while others ask for information for all jobs available (third and fourth jobs, etc.). As a result, details concerning income sources and labor time can vary considerably.

In order to minimize these differences, the variables that are given in each job are used first. Then, when necessary, the time variables that are needed to determine the employment time classifications are created. Once all jobs have the same time variables, these can be analyzed consistently. In terms of the missing information regarding third and fourth jobs, there is little to be done. Moreover, the lack of information in third and fourth jobs, such as labor time questions, makes it impossible to determine accurately the amount of time the individual spent at that job. However, in order to address this issue, when the returns from one of the main or secondary jobs appear to be similar for a third of fourth job, then the median labor time estimates can be used to determine the missing information. (To see in which countries we applied this framework, please

[8] This is a rather generous assumption, intended to minimize the number of observations that are changed and to allow for the instances when individuals work extraordinary numbers of hours in short periods. Both hours per day and hours per week assumptions allow so that no more than a handful of values are replaced.

[9] Given that some surveys do not explicitly differentiate between main and secondary jobs, but instead refer to first and second jobs (or third and more), it can be difficult to definitively confirm that first job listed is in fact an individual's main job. Researchers can use the available data to determine which job they considered to be the main job, applying criteria such as profitability, earnings, labor time, or others that seem appropriate.

refer to Appendix II). Finally, a variable to categorize all the jobs an individual reported in the survey has been created accordingly.[10]

2.3.3 Period of work

It is typical for wage employment questions (participation, labor time, income, and so forth) to be asked for a specific time period, i.e. "the last 7 days (or week)" or "the last 12 months". In some surveys all labor questions refer to the same time period, however, in some cases there is a lack of consistency. This creates situations where it is not always possible to perfectly estimate labor time variables. Another challenge can be found when wage information is reported and time information is missing or zero. This occurs when a person doesn't work during the last 7 days but reports earned income during the last 12 months. In these tricky instances, all of the available variables are used to ensure that the estimates are sound. In addition, the estimates are compared for primary jobs and secondary jobs to identify differences, similarities, and to ensure that these are reasonable. Specific information concerning the difficulties encountered when creating time variables in each survey, as well as the manner in which they were resolved, are discussed in Appendix II.

2.3.4 Insufficient Hours per Week Information

The absence of adequate information regarding hours per week or days per week is very rare. However, there are instances when this information is simply not asked for an entire section of a module, such as a third job (but never a first or second job). Since it is not desirable to disregard labor information for any job, a value of hours per week for these observations is approximated.

To do this, the first step is to compare the means and medians of job 3 monthly earned income with those for job 1 and job 2. If there is a great deal of similarity between job 3 and either of the other two, it is possible to make assumptions about the labor time characteristics of job 3 (generally, there have always been a sensible match). For instance, if the means and medians of job 3 are close to those in job 1, one can assume that job 3 also primarily represents other main jobs in the last 12 months. According to each occupation code or industry (depending on what is available in each survey), it is then possible to assign hours per week values to the observations in job 3 based on those in job 1. Generally, job 3 observations are limited to employment that were not previously mentioned in the survey, but were worked in the last 12 months. Often, this accounts for main jobs that were not worked in the last seven days because main and secondary job questions specify this recent time period. Thus, it is likely that most of this residual job section is in fact made up of primary jobs, which are often worked more intensively than secondary jobs and provide more income. That being said, job 3 may also refer to some part time or secondary jobs, In the case that job 3 appears to be similar to a secondary job, the values of hours per week of the secondary job applies. Overall, the rule is to apply the values of hours per week of the job that is most similar in earned income. Though this approach may overestimate labor time in those limited cases, it is still deemed as preferable to completely imputing labor time information. This is due to the fact that predicting these values may drive analytic results, which can create doubts about findings. However, this is not a concerning issue given the few

[10] This variable (JOB) follows the organization (or logic) of each survey and considers the first job surveyed as job one, the second job queried as job two, and so forth, regardless of labor time or earnings considerations that may create complications. This approach is applied because it is the simplest way to organize the employments consistently across numerous surveys.

cases (and the few observations) in which we apply this measure. Please refer to Appendix II for more country-specific explanations regarding the insufficient hours per week information.

3. Daily wages and monthly earnings

Having categorized wage employment based on labor time characteristics, the methodology for determining earnings and wages for labor participants is now presented. Wages are generally assessed over as short a time period as possible to calculate the return to labor over that time period. Earnings are generally more similar to income and are used to assess over monetary gains from participating in labor markets over a longer period.

Defining both wages and earnings requires considering the time units reported in each survey. This entails both the time units for labor time participation, i.e. days worked per month or per year, as well as the time units for returns, i.e. compensation received per day, per month or per year. Ideally, sufficient information is available to calculate wages and earnings over multiple periods (per hour, day week, month or year) and wages over different periods so that different units can be used for comparisons. However, given the multi-country nature of the RIGA-L database, creating comparable wages is complicated by the fact there is variation in the way questions are asked. In the end, for reasons described below wages are presented using daily wages and earning at a monthly level. The following part provides a more detailed explanation of how this is done and the assumptions employed to do so.

3.1 General Principles

For the purpose of this study, which focuses solely on labor markets, an employment income aggregate for the individual is created taking into account the different sources of labor income. As noted by Carletto, et al., (2007, p. 3) employment income is made up of "…all income received in the form of employee compensation either in cash or in kind." In each survey, sources of labor income earned vary depending on the country and nature of the rural economy. As such, employment modules generally ask two types of remuneration questions: Cash and In-Kind. Cash questions are related to income that is earned as a wage, salary, or tips while in-kind questions usually refers to payments in the form of food, clothes, livestock, transportation, housing, and so forth.. In most of the surveys, values for in-kind income are provided; however, in the case when values are inexistent prices are calculated for the relevant products using data from the survey's consumption module. These prices are then applied to the quantities of the in-kind products reported in the employment module to estimate their equivalent value of earned income.

In order to create employment income measures that are comparable across countries and over time, the following criteria are applied in the estimation of income measures:

- For each survey, only the *rural sample* is focused on.[11]

[11] Given the motivations of RIGA, the construction of the RIGA-L database is motivated by a desire to better understand the *rural* labor market. That being said, all of this data is also created for urban observations and is

- All income is calculated at both the *job and individual level*. This allows identifying the amount of income earned for each job, as well as the total for an individual with more than one job.
- All income earned is estimated as *monthly.*
- All wages are measured as *daily*.
- All income components are *net of costs*.[12]
- All income is reported in *local currency units*.
- All income is categorized by *industry*.

Income earned is estimated on a monthly basis, as opposed to annually, because monthly is the most common time period for income questions in employment modules. This is especially the case for inquiries that are asked in cash, as opposed to in-kind, as well as those for first jobs, as opposed to additional employments (second, third, and so forth). As a result, relying on monthly is the most convenient option available and should also be the most accurate, since earned income is estimated in the same time period that respondents report it. In addition, this approach is computationally simpler, and possibly sounder, because fewer assumptions and conversions are necessary. Nonetheless, some income questions are asked for in hourly, daily, weekly, two week, 15 day, half-month, or annual time periods, and must be converted to monthly using the labor time questions available. Even when all these options are available in the survey, it is found that most respondents report monthly periods. If, for instance, either hourly or yearly is the time period chosen, a wider range of conversions (often relying on more assumptions when the requisite labor time variable are lacking) are necessary. As noted earlier, this is a constant concern because most surveys only inquire about a handful of labor time units and it is necessary to ensure comparability over numerous surveys from different countries for the RIGA-L data.

However, monthly income does not provide the best possible wage estimate since there can be great variation in the amount of time worked in a month and thus this does not accurately reflect the return to labor. To calculate wages, it is preferable to consider the amount of employment income earned per a much smaller time unit, such as week, day or hour. In the RIGA-L database the standard wage estimate is income earned per day for reasons similar for choosing monthly as the standard period for income earned. First of all, days worked per month are a more common labor time measure than hours per day or week, as well as weeks per month or year. In addition, converting income earned from months to days avoids an additional step that would be necessary for conversions to hourly wages.[13] The manner in which days per month is calculated will be discussed in the following section, along with the practicalities of estimating monthly income earned and daily wages.

---

referred to as the Urban Income Generating Activities Labor initiative (UIGA-L). When using the labor data simply search for the urban variable (URBAN) and specify the group of interest.

[12] Taxes, such as social security, is the only cost that has been subtracted from gross income earned to create net income earned.

[13] In the case that the required labor time variables are not available, such as hours per day or week, additional assumptions (that may not reflect the reality on the ground or distort calculations) would be necessary to make the measures comparable.

3.2 Applying the Framework

In order to ensure comparability across countries, a consistent framework has been adhered to when creating monthly earned income and daily wage variables. This approach aims to estimate income information in the simplest and most accurate manner, with precedence always being given to income information as it has been reported in the survey. As such, assumptions and conversions are only applied when no other reasonable options exist. That being said, exceptions to this methodology do occur because of country specific situations or a lack of sufficient income or labor time information in a survey.

The first step of this process entails identifying what questions in the employment module refer to employment income earned, as well as what time period these refer to. The questions that are asked on a monthly basis require no additional computation and are transformed into variables immediately. However, it is often the case that income questions refer to a different time period, such as per day or per year, amongst others. In these cases, the existing labor time questions for each survey are employed to convert this information into monthly income earned. For instance, if a question about tips is reported annually and a question about number of months worked in the last year exists, then a "monthly tips earned" variable can be created simply by dividing annual tips by the number of months worked. The following table summarizes the method for converting income earned with existing labor time information:

Table 3. Methods for Converting Earned Income to Monthly Values

Reported Income	Conversions
Annual	- Divide by the reported months worked in a year
Semester/Half Year	- Divide by the reported months worked in the semester/half year
Trimester	- Divide by the reported months worked in the trimester
15 Days/Half Month	- Divide by the average of the reported days worked per month divided by 15.
14 Days	- Multiply by the average of the reported days worked per month by 14.
Weekly	- Multiply by the reported weeks worked per month
Daily	- Multiply by the reported days worked per month; or multiply by the reported days worked per week times weeks worked per month
Hourly	- Multiply by the reported number of hours worked per day times days worked per week, times weeks worked per month; or multiply by the reported number of hours worked per week times weeks worked per month

Unfortunately, it is sometimes the case that there is insufficient information to convert income earned that was reported for a time period other than monthly. When such a situation occurs, conversions are made based on assumptions for the amount of time worked, similar to the way labor time variables are previously estimated. It should be noted, that the number of observations

this affects is generally quite small.[14] In addition to the labor time assumptions that have already been explained above, the assumption regarding days worked per month is 30.4375 (365.25/12). This is used instead of 31 days per month, which is the maximum number of days per month. Though the magnitude of difference between these two values is not large, 30.4375 is employed because conceptually it is a considered a more precise estimate than the maximum number of days per month.

Once all of the monthly income earned variables have been created, the next step is to check for outliers. This is an important procedure and, as such, a full section below has been dedicated to how outliers are dealt with. After the first outlier check, it is then possible to existing monthly earned income variables (aggregated according to the categories discussed later in section 3.3.2) into one variable for total monthly earned income (WGE_M). During this aggregation, costs are also taken into account to ensure that the final variable is net of costs, as opposed to gross (which could overestimate the income an individual actually has at his or her disposal). So far, the only reported cost, which is subtracted during the aggregation process, has been income tax (i.e., the contribution to social security and health system). Once the variable is aggregated monthly earned income undergoes a second outlier check, which will be discussed later, before being considered final.

Having completed the monthly earned income estimation, it is possible to create a daily wage. Simply put, this is achieved by dividing monthly earned income by the number of days per month worked for each observation. Consequently, a variable for days per month must be created. In many cases, a question about days per month exists in the employment modules, which makes this process very straightforward. Nonetheless, in a limited number of cases days per month must be created by converting other remaining work time information provided by a survey or, as a last resort, days per month must be estimated based on assumptions similar to those described previously. Again, it should be noted that in most cases values for days per month are created for very few observations based on assumptions, as this is one of the most frequently reported time periods in employment modules. When other work time variables are used to calculate days per month, the following approach is applied:

Table 4. Labor Time Conversions

Labor Time Reported	Conversions
Weeks per month	- Divide by the reported number of days worked per week.
Days per year	- Divide by the number of reported months worked (per year).
Days per week	- Multiply by the number of reported weeks worked per month.

If there is a unique case when assumptions must be relied on to estimate days per month, those that have already been listed here and in the labor section above are applied.[15] Having created reliable days per month estimates for labor participants, it is then possible to create a daily wage based on the monthly wage. Daily wage is finally created by dividing the aggregate monthly income by the number of working days per month.

[14] More often than not, insufficient labor time information is found for secondary or other employments, not main jobs.

[15] For any country or survey specific discrepancies see Appendix II.

All monthly income earned and daily wage information is classified by industry in a consistent fashion. Similar to the approach that Carletto, et al., (2007, p. 3), all labor employment data is disaggregated by industry across countries. The disaggregation is based on the United Nations' International Standards Industrial Classification of All Economic Activities (ISIC).[16] Initially, employments are grouped into ten principal industry categories: (1) Agriculture, Forestry and Fishing; (2) Mining; (3) Manufacturing; (4) Utilities; (5) Construction; (6) Commerce; (7) Transportation, Communications and Storage; (8) Finance and Real Estate; (9) Services; and (10) Miscellaneous. Once monthly earnings have been estimated and outlier checks have been completed, broad ISIC categories are combined so as to avoid small sample sizes by industry, making an effort to combine conceptually similar, and similarly remunerated, industries. In the end, the following seven industry categories are created: (1) Agriculture, Forestry and Fishing; (2) Manufacturing; (3) Construction; (4) Commerce, Transportation, Communications, Storage, Finance, and Real Estate; (5) Services; (6) Mining & Utilities; and (7) Miscellaneous. Having divided up the jobs into these categories, it is also easy to compare between agricultural labor (group 1) and non-agricultural labor (the remainder of wage employment, i.e. the aggregate of industry groups 2 through 7).

Lastly, occupation classifications are also developed for all employment activities (based on work activities or the type of job worked) in a consistent manner across surveys. This categorization is based on the International Standard Classifications of Occupations (ISCO) nomenclature provided by the International Labor Organization's (ILO).[17] This is because nearly all of the surveys in the RIGA-L database utilize the ISCO-88 classification system, or one of its predecessors. The occupations for each survey are aggregated into the following ten *Major Groups*: (1) Legislators, Senior Officials, and Managers; (2) Professionals; (3) Technicians and Associate Professionals; (4) Clerks; (5) Service Workers and Shop and Market Sales Workers; (6) Skilled Agricultural and Fishery Workers; (7) Craft and Related Trade Workers; (8) Plant and Machine Operators and Assemblers; (9) Elementary Occupations; (10) Armed Forces Occupations; and (11) Other/Unknown. In practice, it should be noted that jobs in each group are not always found in all the surveys.

3.3 Data Issues

3.3.1 Missing Values & Outliers

At this point of the process, only the participants in the rural employment should be present in the data. However, there is always a possibility of finding a small number of observations that claim to have participated and have missing values in income variables. This can be a result of non-responses by survey participants or survey skip patterns, as well as surveying and data cleaning errors. Generally, only a handful of such observations are found; however, in the case that they are present we first check to see if these observations truly appear to be participants or

[16] The classification system can be found at http://unstats.un.org/unsd/cr/family1.asp. It should be noted that as the world changes the industry classification framework continues to be revised by the United Nations. Consequently, the year of each survey is matched with the nearest ISIC classification standards in order to find the most suitable and applicable revision of industry categories.

[17] For more information please see the following: http://www.ilo.org/public/english/bureau/stat/isco/index.htm

not. For instance, an observation may report a missing value for some income questions, but not others. If that very same observation reports labor time worked, then it is clearly a true participant and the missing values for other income questions can be recoded as zero. On the other hand, if an observation reports missing values for all income and labor time questions, then it can be dropped from the participant sample.

In the case where there are observations that report missing values for all income questions, but report full or partial labor time information, efforts are made to reconcile this discrepancy. If this oddity appears justifiable and it looks like these observations really are participants income is estimated in some manner. Though crude, one option relied upon is to replace the missing values with median earned income values, which is better than simply excluding the observations in question. However, if their appears to be no rhyme or reason to these observations, the missing income values are recoded to zero, which leaves them in the data but essentially defines them as not participating because of a lack of income. This is a rare situation that is not present throughout the surveys and affects less than one percent of observations when present.

The outlier checks used in this analysis are based on those recommended by Carletto, et al., (2007, p. 7). Outlier checks are performed by dividing a monthly earned income variable according to one relevant subgroup. In this case, the logical sorting variable is the industry classification for employment. When a logical sorting variable does not exist or there are an insufficient number of observations in each sorting category, an administrative variable is substituted.[18] Carletto, et al., (2007, p. 7) define an outlier as "…values greater or less than three standard deviations from the median value of the variable for that specific group."

The RIGA-L is also based on this definition, but adds two amendments to the outlier check process they recommend in order to tailor the procedure labor data. First of all, outlier checks are weighted according to the individual weights provided by each survey in order to take into account how representative each observation is of the overall population. Secondly, monthly income earned variables are substituted with the log of original monthly income variables. This revision is included because if this procedure does not take into account logs, it exhibits a bias towards classifying high values (the right tail of a distribution) as outliers while ignoring low values (the left tail of a distribution) that may be just as dubious. This can be especially problematic for labor data because outliers are also likely to be found in the left tail of an employment income distribution. This is because the right tail of a labor income distribution is generally quite long. If this measure is not implemented, the bias described above is likely intensified and may influence results.

As Carletto, et al., (2007, p. 7) suggest: "…zeroes and missing values are excluded from the computation of the median, standard deviation and identification of outliers in order to achieve accurate imputations. This ensures the medians and standard deviations are not skewed by zeros and that households with missing values are not erroneously assigned values." The same approach (with the incorporated RIGA-L changes) is also followed here.

[18] Outlier checks should be performed for groups of 50 or more observations to ensure that outliers are accurately identified. This criterion is adhered to in most cases; however, in some instances when there are few observations participating in labor employment, outlier check sub-groups may be slightly smaller.

An initial outlier check is implemented after raw data has been transformed into monthly earned income variables. Outlier values are flagged and replaced with the median value of monthly earned income for each variable. Given the approach outlined above, less than one percent of observations are normally affected by this check. After the first outlier check takes place and monthly income earned variables are aggregated, the revised monthly earned income variables undergo a second outlier check. This time around, the same process is applied to the corrected variable using an administrative or geographic variable for the purposes of grouping. Once again, less than one percent of observations are regularly flagged as outliers and replaced with median values. This approach is applied throughout the RIGA-L data construction process in order to systematically minimize the number of observations and values altered.

3.3.2 General Issues

When it comes to wages there are three additional challenges to be considered in order to make comparable income variables across countries. The first challenge is that income questions are often asked in an assortment ways in each survey. On one hand, there are surveys that provide extremely detailed information about income, such as questions about tips, bonuses, vacations, or social security. Similarly, some surveys contain detailed questions that disaggregate types of in-kind compensation, such as food, clothing, housing, transportation, and so forth. On the other hand, there are surveys that provide only a single aggregated question about different types of income earned. The second challenge is that the income earned questions also refer to a variety of payment periods. In some cases, surveys ask about income for one consistent time unit, such as monthly wages; while others provide numerous options to report income in, such as daily, weekly, every two weeks, half month, monthly, etc. This challenge can exist either across different income questions, or sometimes within the very same query where respondents are given the option to self-specify the appropriate payment period.

A third challenge is related to the fact that income questions, and the detail they are asked in, different within an employment module according the type of job. For instance, sometimes the level of detail asked about decreases as you move from job 1 to job 2 and job 3; especially if the progression in the employment module is form the main job, to a secondary employment, and finally an extra job. In general, questions tend to be the same for the first two jobs (if there are multiple), but if there is a third job listed, fewer questions are asked or they are asked differently. This can result in a lack of both information and consistency for earned income information across types of jobs and portions of the employment module. A final challenge can be found when all three of the issues above exist within the same survey. Though instances of this are unlikely, overcoming these issues is a continual task.

To deal with these challenges and simplify the process of estimating monthly earned income as much as possible every effort is made to aggregate income wages into four consistent categories, depending on the information provided in the surveys. Theoretically, the four variables categories are:

Table 5. Income Earned Variables

Cash	In-Kind	Cost	Bonus
1. Cash	1. Livestock	1. Social Security	1. Miscellaneous
2. Salary	2. Equipment	2.Other Taxes	2. End-of-year
3. Wage	3. Clothes		3. Vacation
5. Tips	4. Housing		
	5. Food		
	6. Transport		

Later on, all of these variables are aggregated, depending on what is available, to create a net monthly earned income variable. In other words, Cash, In-Kind, and Bonus are added together while Cost is subtracted from the total. Lastly, as explained above, most cash, salary, wage, or tips income earned questions are asked in terms of months, but if questions are expressed in terms of other time periods, these are converted to months by using the labor time variables available prior to aggregating them. This method is usefully because it minimizes the need to make labor time assumptions; however, in some cases they are necessary depending on the information available in each survey for a specific job.

3.3.3 Insufficient Days per Month Information

In order to create daily wages it is necessary to have information regarding days worked per month. However, as explained in sections 2.3.2 and 2.3.4, there are cases where a considerable amount of labor time information is missing. This is particularly true when a third job is reported and fewer questions concerning labor time variables or income are asked for a residual job. When this occurs for days per month, the variable is created following the same methodology described in section 2.3.4.

4. Individual Characteristics

In addition to household characteristics previously created to accompany the RIGA (household level income aggregate) database, a few individual level variables are created in order to facilitate more robust individual level analysis. The variables are limited to a handful of human capital characteristics, such as gender, age, and years of education, when available. The manner in which these variables are created is straightforward and, as a result, will not be discussed. A list of the variables created in the human capital dataset and in the individual characteristics dataset is found in Appendix I.

5. Using RIGA-Labor (RIGA-L) Data

5.1 Final Data

The final data created for each country can be found in the following datasets:

Wages:
-EmploymentAggregate-Jobs.dta
-EmploymentAggregate-Individuals.dta

Individual Characteristics:[19]
-CountryYear_hc.dta (i.e., Tajik03_hc.dta)
-CountryYear_indchar.dta (i.e., Tajik03_indchar.dta)

It should be noted that the employment data is available in two options: (1) employment data at the job level; and (2) employment data at the individual level. To be more specific, there is one observation for each job worked (regardless of if they person working that specific employment has more than one job) in the EmploymentAggregate-Jobs database. On the other hand, there is one observation for each person in the EmploymentAggregate-Individuals database.[20] CountryYear_hc.dta and and CountryYear_indchar.dta match each of these data sets.

Datasets should be merged by using the household identifier (HH) and the individual identifier (INDID).[21] In the job level employment aggregate an additional variable (Job) is included for unique identification. This makes it possible to combine employment data, such labor time or wages, with individual characteristics, such as gender or years of education, and analyze them together. As mentioned previously, a list of the variables created can be found in Appendix I, as well as a description of each variable and the unit in which it is created.

5.2 Unique Identification and Linking to the RIGA Household Data

Standard variables are created for all countries to uniquely identify households (HH), individuals (INDID), and jobs (JOB).[22] In the majority of cases, unique household and individual identification variables are already available in the raw survey data and are merely renamed for consistency. However, in some cases it is necessary to create a unique household identifier by combining numerous variables, such as the primary sampling unit (PSU) or region (REGION) – or even a combination of variables such as city, area, zone, etc. – in combination with the non-unique household identifier (HHID per the relevant sub-group). However, the original survey's household identifier (HHID) has been kept in the case that researchers need to merge RIGA-L with data from the original country surveys. Details for atypical situations that merit more explanation can be found in Appendix II. Lastly, all individual labor datasets can be integrated

[19] Note: Country and year are abbreviated.
[20] This means that if a particular person works more than one job the multiple observations at the job level has been collapsed into one at the individual level. In the case of labor time, income earned, or wages, values of are summed during the collapse procedure to reflect all of the time worked or income earned by each person.
[21] Only one country dataset does not follow this rule, Malawi 2004 (see Appendix II for more details).
[22] It should be noted that the creation of household identifiers (HH) follows the same approach and syntax as those applied by the RIGA household income aggregate project.

with the existing household RIGA household data by merging the relevant databases with the household identifier (HH), which is consistent across databases.[23]

[23] Given that the individual RIGA-L database applies added criteria (in addition to what is applied for the RIGA household income aggregate database) to determine participation in wage employment, there is a slight difference in the number of participants between the two. This is due to the fact that all lacking both income and labor time information are dropped in the individual analysis, even if they claimed to have participated in wage employment, while they are not necessarily removed form the RIGA household data.

References

Carletto, G., Covarrubias, K., Davis, B., Krauzova, M., and Winters, P. (2007). Rural Income Generating Activities (RIGA) Study: Income Aggregate Methodology. Agricultural Sector in Economic Development Service, Food and Agriculture Organizations of the United Nations (FAO).

Appendix I

1. Surveys and Participation in Wage Employment

The RIGA-L data is created for a total of 15 surveys from different countries ranging from 1995 to 2005. Below is a table listing each survey as well as the number of household and individual labor market participants in the urban and rural areas.

Table 1. Survey Details

Country	Survey	Households		Individuals	
Sub-Saharan Africa		*Urban*	*Rural*	*Urban*	*Rural*
Ghana98	Ghana Living Standards Survey Round 3	714	661	806	757
Malawi04	Integrated Household Survey - 2	1,044	5,953	1,489	9,686
Nigeria04	Living Standards Survey	1,327	1,560	1,622	1,855
South & East Asia					
Bangladesh00	Household Income-Expenditure Survey	1,601	3,010	2,254	4,058
Indonesia00	Family Life Survey - Wave 3	3,172	2,504	4,926	3,593
Nepal03	Living Standards Survey II	684	1,602	1,156	3,051
Vietnam98	Living Standards Survey	1,080	1,862	2,115	3,417
Eastern Europe & Central Asia					
Albania05	Living Standards Measurement Survey	1,182	517	1,719	629
Bulgaria01	Integrated Household Survey	872	248	2,086	643
Tajikistan03	Living Standards Survey	869	1,697	1,209	3,215
Latin America					
Ecuador95	Estudio de Condiciones de Vida	2,348	1,456	4,414	2,724
Guatemala00	Encuesta de Condiciones de Vida	2,509	2,525	4,754	4,425
Nicaragua98	Encuesta de Medicion de Niveles de Vida	1,573	1,032	2,781	1,823
Nicaragua01	Encuesta de Medición de Niveles de Vida	1,735	1,096	3,184	1,928
Panama03	Encuesta de Niveles de Vida	2,558	1,776	4,491	2,956

Notes: (1) Participants are only those of working age (15 to 60 years old). (2) Households may have more than one participant in wage employment. (3) Urban employment is only for the non-agricultural sector. (4) Rural employment in Malawi is predominantly Ganyu labor.

2. Variables

Below is a list of the employment aggregate variables created for each survey:

Table 2a. Employment Variables

Output Data Files	Variables	Unit	Description
Countryyear_IND_WGEJOB.dta			
Administrative			
	hh	Household	Household Identifier
	indid	Individual	Individual Identifier
	job	Job	Indicates if job is first, second, third, etc.
	indweight	Individual	Individual Weight
Job			
	job1	Job	Indicates if job is first job (first job==1)
	job2	Job	Indicates if job is second job (secondary job ==1)
	job3 (etc.)	Job	Indicates if job is third job (if available, third job ==1)
	occupation	Job	ISCO-88 Major Occupation Code
	public	Job	Indicates if Job is in the Public Sector (==1)
Labor Time			
	fyft	Job	Indicates if the job is full year and full time (==1)
	fypt	Job	Indicates if the job is full year and part time (==1)
	pyft	Job	Indicates if the job is part year and full time (==1)
	pypt	Job	Indicates if the job is part year and part time (==1)
	tot_months	Job	Total months worked
	agr_months	Job	Months worked in agriculture
	non_agr_months	Job	Months worked in non-agricultural activities
	months1	Job	Months worked in industry 1
	months2	Job	Months worked in industry 2
	months3	Job	Months worked in industry 3
	months4	Job	Months worked in industry 4
	months5	Job	Months worked in industry 5
	months6	Job	Months worked in industry 6
	months7	Job	Months worked in industry 7
	months8	Job	Months worked in industry 8
	months9	Job	Months worked in industry 9
	months10	Job	Months worked in industry 10
	months11	Job	Months worked in industries 6, 7 and 8
	months12	Job	Months worked in industries 2 and 4
	tot_hrsweek	Job	Total hours worked per week
	agr_hrsweek	Job	Hours worked per week in agriculture
	non_agr_hrsweek	Job	Hours worked per week in non-agricultural activities
	hrsweek1	Job	Hours per week worked in industry 1
	hrsweek2	Job	Hours per week worked in industry 2
	hrsweek3	Job	Hours per week worked in industry 3

	hrsweek4	Job	Hours per week worked in industry 4
	hrsweek5	Job	Hours per week worked in industry 5
	hrsweek6	Job	Hours per week worked in industry 6
	hrsweek7	Job	Hours per week worked in industry 7
	hrsweek8	Job	Hours per week worked in industry 8
	hrsweek9	Job	Hours per week worked in industry 9
	hrsweek10	Job	Hours per week worked in industry 10
	hrsweek11	Job	Hours per week worked in industries 6, 7, and 8
	hrsweek12	Job	Hours per week worked in industries 2 and 4
	tot_daysmonth	Job	Total days worked per month
	agr_daysmonth	Job	Days per month worked in agriculture
	non_agr_daysmonth	Job	Days per month worked in non-agriculture
	daysmonth1	Job	Days per month worked in industry 1
	daysmonth2	Job	Days per month worked in industry 2
	daysmonth3	Job	Days per month worked in industry 3
	daysmonth4	Job	Days per month worked in industry 4
	daysmonth5	Job	Days per month worked in industry 5
	daysmonth6	Job	Days per month worked in industry 6
	daysmonth7	Job	Days per month worked in industry 7
	daysmonth8	Job	Days per month worked in industry 8
	daysmonth9	Job	Days per month worked in industry 9
	daysmonth10	Job	Days per month worked in industry 10
	daysmonth11	Job	Days per month worked in industries 6, 7 and 8
	daysmonth12	Job	Days per month worked in industries 2 and 4
Wages			
	tot_wge_m	Job	Total monthly income
	agr_wge_m	Job	Agricultural monthly income
	non_agr_wge_m	Job	Non-Agricultural monthly income
	wge_m1	Job	Monthly income in industry 1
	wge_m2	Job	Monthly income in industry 2
	wge_m3	Job	Monthly income in industry 3
	wge_m4	Job	Monthly income in industry 4
	wge_m5	Job	Monthly income in industry 5
	wge_m6	Job	Monthly income in industry 6
	wge_m7	Job	Monthly income in industry 7
	wge_m8	Job	Monthly income in industry 8
	wge_m9	Job	Monthly income in industry 9
	wge_m10	Job	Monthly income in industry 10
	wge_mimp1	Job	Final Imputed: monthly income in industry 1
	wge_mimp2	Job	Final Imputed: monthly income in industry 2
	wge_mimp3	Job	Final Imputed: monthly income in industry 3
	wge_mimp4	Job	Final Imputed: monthly income in industry 4
	wge_mimp5	Job	Final Imputed: monthly income in industry 5
	wge_mimp6	Job	Final Imputed: monthly income in industry 6
	wge_mimp7	Job	Final Imputed: monthly income in industry 7

	wge_mimp8	Job	Final Imputed: monthly income in industry 8
	wge_mimp9	Job	Final Imputed: monthly income in industry 9
	wge_mimp10	Job	Final Imputed: monthly income in industry 10
	wge_m11	Job	Final Imputed: Monthly Income for industries 6, 7 and 8
	wge_m12	Job	Final Imputed: Monthly Income for industries 2 and 4
	tot_wge_d	Job	Total daily wage
	agr_wge_d	Job	Agricultural daily wage
	non_agr_wge_d	Job	Non-Agricultural daily wage
	wge_d1	Job	Daily wage in industry 1
	wge_d2	Job	Daily wage in industry 2
	wge_d3	Job	Daily wage in industry 3
	wge_d4	Job	Daily wage in industry 4
	wge_d5	Job	Daily wage in industry 5
	wge_d6	Job	Daily wage in industry 6
	wge_d7	Job	Daily wage in industry 7
	wge_d8	Job	Daily wage in industry 8
	wge_d9	Job	Daily wage in industry 9
	wge_d10	Job	Daily wage in industry 10
	wge_d11	Job	Daily Wage for industries 6, 7 and 8
	wge_d12	Job	Daily Wage for industries 2 and 4
Participation			
	p_tot_wge_m	Job	Participation in wage employment (participant ==1)
	p_agr_wge_m	Job	Participation in agricultural wage employment (participant ==1)
	p_non_agr_wge_m	Job	Participation in non-agricultural wage employment (participant ==1)
	p_wge_m1	Job	Participation in industry 1 wage employment (participant==1)
	p_wge_m2	Job	Participation in industry 2 wage employment (participant==1)
	p_wge_m3	Job	Participation in industry 3 wage employment (participant==1)
	p_wge_m4	Job	Participation in industry 4 wage employment (participant==1)
	p_wge_m5	Job	Participation in industry 5 wage employment (participant==1)
	p_wge_m6	Job	Participation in industry 6 wage employment (participant==1)
	p_wge_m7	Job	Participation in industry 7 wage employment (participant==1)
	p_wge_m8	Job	Participation in industry 8 wage employment (participant==1)
	p_wge_m9	Job	Participation in industry 9 wage employment (participant==1)
	p_wge_m10	Job	Participation in industry 10 wage employment (participant==1)
Countryyear_IND_WGEIND.dta			

Includes all variables in IND_WGEJOB.dta (shown above) and the numjobs variable. This is the same dataset as IND_WGEJOB.dta but collapsed at the individual level.			
Job			
	numjobs	Individual	Indicates the number of jobs of an individual

Below is a list of the population sample variables created for each survey:

Table 2b. Sample Variables

Output Data Files	Variables	Unit	Description
Countryyear_IND_ADMIN.dta	hh	Household	Household Identifier
	indid	Individual	Individual Identifier
	original household ID	Household	Original Household Identifier (Raw Data)

	original individual ID	Individual	Original Individual Identifier (Raw Data)
	urban	Household	Location (Urban =1; Rural = 0)
	indweight	Individual	Population weight factor
	quintile	Household	Expenditure Quintiles - Rural
	quinturb	Household	Expenditure Quintiles - Urban
	decile	Household	Expenditure Deciles – Rural
	decilurb	Household	Expenditure Deciles - Urban
	pcexp	Household	Per-capita Expenditure
	region/ division	Household	Indicates the administrative division of the household.

Below is a list of the individual characteristic variables created for each survey, which accompany the RIGA household characteristics:

Table 2c. Individual Characteristics Variables

Output Data Files	Output Variables	Unit	Description
Countryyear_HC_CHAR..dta	gender	Individual	Gender of individual. Male =1 and female = 2
	rel (or relation)	Individual	Relationship with head of household
	age	Individual	Age in years
	indlabort	Individual	Indicates if the individual is within the working age group (between 15 and 60) (==1)
	mlabort	Individual	Indicates if the individual is within the male working age group (between 15 and 60) (==1)
	flabort	Individual	Indicates if the individual is within the female working age group (between 15 and 60) (==1)
	edu	Individual	Number of years of education
	religion	Individual	Religion of the individual (Not always available)
	Other ethnicity status variables (nativel, indigen)	Individual	Indicates if the individual is indigenous, or other. (This variable is not available in all countries).

Appendix II

1. Country Specific Issues

Deviations from the methodology are sometimes necessary due to survey specific issues. A list of these variations is briefly presented below:

1.1. Albania 2005

No amendments to the methodology required.

1.2. Bangladesh 2000

No amendments to the methodology required.

1.3. Bulgaria 2001

Occupation codes in this survey are not based on ISCO nomenclature. A unique set of codes (found on page 57 of the questionnaire) is used. These categorizations are manually converted to fit unto the ISCO-88 classification system (previously described in section 3.2 of the methodology) in order to facilitate consistent cross country analysis.

1.4. Ecuador 1995

In this survey, information on days worked per week is missing for job 3 (main job in the last 12 months not worked in the last week). This information is needed to calculate the number of days per month worked (by multiplying days worked per week by 4.35). In order to obtain days per week (DAYSWEEK), we divide hours worked per day (HOURSDAY) by 8, assuming that 8 hours of work equal 1 day of work. When doing this one produces days per week results greater than 7 days of work around 16% of the observations; these are recoded with the maximum days of worked possible in a week (7). It should be noted that by assuming 8 hours worked per day the median for days worked per week in job 3 turns out to be 5, which is similar to the results of job 1 (main job in the last 7 days). Even though this approach is not the ideal, it is the best possible approach given the available information. The same approach is followed for job 4 (secondary job in the last 12 months not worked in the last week), which also lacks days per week information.

Additionally, as in Bulgaria 2001, occupation codes in this survey are not based on the ISCO nomenclature. Instead, the Ecuador survey relies on the Codigo Industrial Internacional Uniforme – Revision 3 (CIIU-3), which can be found in the accompanying activity codes documentation. Again, these categories are manually recoded to fit the ISCO-88 major groups (as detailed in section 3.2 of the methodology).

1.5. Ghana 1998

In this survey, information concerning the duration of labor in all jobs is missing. In the household RIGA database, duration of 12 months worked in the last year is, by default, assumed. Consequently, this approach is also followed to create the labor time categories here. However, this method is not ideal because it may overestimate the duration of employments. This being said, the lack of information prevents determining the precise duration of each job (full year or part year) and for this reason, only frequency (full time and part time jobs) are differentiated. Subsequently, the four labor time categories are converted (in analysis files) in order to rely on the labor time information available (it would be inaccurate to analyze FY or PY groups). This is to say that all the observations are classified as full-time (FT) or part-time (PT) employments and analyzed accordingly.

Additionally, as in Ecuador 1995, information on days worked per week is missing. We follow the same assumption as in Ecuador 95 (8 hours of work equal 1 day of work) and divide hours worked per week by 8. Once this is done, the median days per week in job 1 is around 4.8, in job 2 it is nearly 4.6, and in job 3 it is roughly 2.6.

1.6. Guatemala 2000

In this survey, the only labor time information provided for job 3 is months worked per year. Consequently, the procedure described earlier in section 2.3.4 of the methodology is applied.

1.7. Indonesia 2000

Again, as in Ghana 1998 and Ecuador 1995, information on days worked per week is missing so 8 hours of work is assumed to equal 1 day of work. As a result, the median of days worked per week is about 5.6 in for job 1 and approximately 2.8 for job 2.

1.8. Malawi 2004

The labor time information available for Ganyu labor (job 2 – casual labor) is days per year and hours per week in the last 7 days. Therefore, the variable weeks per year is created by dividing days per year by 7. In this case, 7 days per week is assumed as opposed to 6 because assuming 6 days created numerous values above 52 weeks. In addition, days per month are created by dividing days per year by 12 (once again, months would have been preferable if available).

It should also be noted that there are 5,836 Ganyu observations (out of approximately 9,000) with 0 hours per week who do report employment income. This is because the hours per week question in this survey only refers to the last 7 days. Consequently, it is assumed that the observations in question worked, but not in the last 7 days. In this case, the median of hours per work for Ganyu labor (excluding observations 0 hours per week) is used to replace the 0 value for the observations in question.

Although this is a crude technique, it is not expected to influence outcomes much. First of all, the median hours per week turn out to be around 12, which is not unreasonable considering the unstable nature of casual Ganyu labor. In addition, if the value of these observations had not been replaced, these observations would have been categorized as part time (< less than 35 hours per week) to start with; in other words, this imputation does not affect where these observations are categorized according to labor time. Lastly, if one investigates the hours per day distribution (excluding the observations in question) it becomes evident that the vast majority of observations in the Ganyu section fall into the part time category (approximately 93 percent).

Additionally, for Ganyu labor the survey only asks about daily salaries. In addition, minimal labor time information is provided (days per year). To covert daily to monthly salaries, the daily variable is multiplied by 15.215 days per month. The assumption is that, since Ganyu is part-time labor, individuals work only half of 30.4375 days per month. Although this is not a foolproof approach, it is the best solution available.

Lastly, it should also be noted that the population sampling unit variable (PSU) should be included to uniquely identify observations in this data. That is, PSU HHID INDID, are the variables required for unique identification and accurate merging.

1.9. Nepal 2003
No information regarding public and private sector was available in the survey. Therefore, there is no *public* dummy variable in the employment-aggregates datasets.

1.10. Nicaragua 1998
No amendments to the methodology required.

1.11. Nicaragua 2001
No amendments to the methodology required.

1.12. Nigeria 2004
As in Ghana 1998, information on the duration of labor time is missing in job 1 (in job 2, the variable weeks is available). In the household RIGA database, a duration of 12 months worked in the last year is, by default, assumed. Subsequently, the same approach is used here to create the labor time categories. However, this method is not ideal because it may overestimate the duration of jobs in Nigeria. This being said, the lack of information prevents the precise determination of duration for job 1 (FY or PY) and for this reason only frequency (FT and PT) classification can be differentiated. Subsequently, the four labor time categories are converted (in analysis files) in order to rely on the labor time information available (it would be inaccurate to analyze FY or PY groups). This is to say that all the observations are classified as full-time (FT) or part-time (PT) employments and analyzed accordingly.

1.13. Panama 2003
As in Guatemala 2000, job 3 lacks all labor time information except for months worked per year. Consequently, the procedure described earlier in section 2.3.4 of the methodology is applied.

1.14. Tajikistan 2003
No amendments to the methodology required.

1.15. Vietnam 1998
No amendments to the methodology required.