

DATA SOURCES FOR THE
ANALYSIS OF LABOUR MARKET
INEQUALITY IN BRAZIL AND INDIA

CEBRAP-IHD Research Project on
Labour Market Inequality in Brazil and India



INSTITUTE FOR HUMAN DEVELOPMENT
www.ihdindia.org



CEBRAP

centro brasileiro de análise e planejamento

WORKING PAPER NO.
WP 03/2014

***CEBRAP-IHD Research Project on Labour
Market Inequality in Brazil and India***

**Data Sources for the Analysis of Labour
Market Inequality in Brazil and India**

Alexandre de Freitas Barbosa, University of São
Paulo and CEBRAP, São Paulo;

Maria Cristina Cacciamali, University of São
Paulo and CEBRAP, São Paulo;

Ashok Pankaj, Council for Social Development, New Delhi
(formerly Institute for Human Development, New Delhi);

Gerry Rodgers, Institute for Human Development, New Delhi;

Vidhya Soundararajan, Institute for Human Development, New Delhi;

Including contributions by:

Fabio Tatei, Rogerio Barbosa and Ian Prates, CEBRAP, São Paulo;

J. Krishnamurty, Institute for Human Development, New Delhi

São Paulo and New Delhi
2014

Published by:

INSTITUTE FOR HUMAN DEVELOPMENT

NIDM Building, IIPA Campus, IP Estate, New Delhi-110002

Phones: +91-11-2335 8166/ 2332 1610 • Fax: +91-11-2376 5410

E-mail: mail@ihdindia.org • Website: www.ihdindia.org

© Institute for Human Development, 2014

ISBN: 978-81-88315-39-0

Subscription Amount: ₹ 50/- /US \$ 10

LABOUR MARKET INEQUALITY IN BRAZIL AND INDIA

**A comparative study, carried out by CEBRAP, São Paulo and
the Institute for Human Development, New Delhi, with support from IDRC**

Project Description

High inequality in income and welfare is a major policy concern in both Brazil and India, for it undermines efforts to reduce poverty and promote inclusive growth. Over the last decade, the connections between inequality and growth, and between inequality and poverty reduction, have been receiving increasing attention in both national and international development communities. There are many sources of income inequality – production structures, the distribution of assets, the relative power of capital and labour, political forces and social hierarchy, as well as differences in education and capability. But among these many factors, labour market structures and institutions are of central importance. Understanding the pattern of labour market inequality and its determinants is therefore essential.

The research project aims to address these issues and their implications for development policies in both Brazil and India. Policy choices in the two countries intersect, but operate in different historical and social contexts, and have had differing degrees of success. Today in particular, the trends in labour market inequality in the two countries are different, and it is important to understand why, how far this results from underlying social and economic institutions and relationships, and how far from policy choices and their implementation. Relying on extensive existing literatures in both countries, but also contributing to these literatures by bringing together historical, macro and micro perspectives, the project aims to add to knowledge and contribute to policy choice through in-depth comparisons of the relationships and outcomes in the two countries.

The methodology of the project combines three different approaches. The first is a long term historical analysis of the social, institutional and economic changes that affect labour market inequality; the second is an empirical analysis of survey data, which investigates the patterns and determinants of inequality and their changes over time; and the third is a process of policy dialogue that brings together social actors and researchers to examine policy implications.

The major points of focus for this project includes gender, race in Brazil and caste in India, regional differences, skill/education and formal – informal employment relationships.

RESEARCH TEAM

Alexandre de Freitas Barbosa, University of São Paulo and CEBRAP, São Paulo

Rogério Barbosa, CEBRAP, São Paulo

Maria Cristina Cacciamali, University of São Paulo and CEBRAP, São Paulo

Taniya Chakrabarty, Institute for Human Development, New Delhi

Nandita Gupta, Institute for Human Development, New Delhi

J. Krishnamurty, Institute for Human Development, New Delhi

Ian Prates, CEBRAP, São Paulo

Gerry Rodgers, Institute for Human Development, New Delhi

Vidhya Soundararajan, Institute for Human Development New Delhi

Fabio Tatei, CEBRAP, São Paulo

CEBRAP-IHD RESEARCH PROJECT ON LABOUR MARKET INEQUALITY IN BRAZIL AND INDIA

Data Sources for the Analysis of Labour Market Inequality in Brazil and India

I. INTRODUCTION

In many parts of the world there has been a growth in economic inequality in recent years, and this has attracted the attention of both researchers and policy makers. High levels of inequality not only undermine efforts to reduce poverty and promote a general increase in living standards; they also raise questions of social justice, of social solidarity, of the role of public policy and of the acceptability of economic systems.

Brazil and India, the two countries considered in this paper, have different histories with respect to economic inequality. Brazil has long been one of the most unequal countries in the world, and remains so today, despite some success in reducing inequality in recent years. Inequality in India, although deeply rooted, was lower than Brazil's in the past, but has been rising rapidly. This difference in trends in inequality between these two countries is of considerable interest for both research and policy. Among the reasons, patterns of work and employment play an important role – differential access to jobs, inclusive or exclusionary labour market institutions, differences in wages and wage shares, patterns of organization and industrial relations. But the relationships are complex and are embedded in the histories and social structures of each country.

This CEBRAP-IHD project on labour market inequality is investigating these issues in India and Brazil in a comparative framework. As a first step, the present paper maps out the main data sources available in the two countries that can be used to analyse the patterns and trends in economic inequality, with a particular focus on the different aspects of labour market inequality. It addresses the following questions:

- (i) What are the main government and other data sources that can be helpful in understanding the nature and sources of economic inequality?
- (ii) What are the characteristics of the data sources in terms of design, coverage and frequency?
- (iii) What is the content of each data source, in terms of measures of inequality and its correlates? How comparable are measures that are repeated over time?
- (iv) What are the analytical issues that can be addressed using each data source?
- (v) How far are the available data on the two countries comparable?

The paper discusses five broad and overlapping categories of data. The first concerns censuses and other data sources that permit a longer term, historical analysis of overall trends. The second concerns regular national surveys on labour markets and employment. The third concerns surveys and censuses on other topics with a bearing on inequality. The fourth covers administrative records. And the fifth briefly identifies some relevant macro-economic data sources.

For each of these categories, we present first the data available for Brazil, then for India, commenting on their availability and the ways they will be or may be used to investigate inequality. We then make some comments on the comparability of the sources for the two countries. A summary is presented in tables 1 and 2 for Brazil and India respectively.

II. HISTORICAL DATA: CENSUSES AND OTHER SOURCES

1. Brazil

(i) *Demographic Censuses and Historical Data*

Demographic Censuses have been carried out every ten years by the Brazilian Institute of Geography and Statistics (IBGE) since 1940. There were censuses in 1960, 1970, 1980, 1991, 2000 and 2010. The entire Brazilian population is interviewed in their homes and answers a multi-themed questionnaire covering demographic and socioeconomic characteristics. A set of limited and summarized questions (around 15 to 20 items) is asked to the entire population for the Demographic Censuses. From 1960, the IBGE introduced a sample survey system that works in conjunction with and parallel to the Census. The population as a whole (also termed *universe* or “*non-sample*”) answers a small number of questions, while a sample completes a more extensive and detailed questionnaire, thus allowing more in-depth socioeconomic investigation. From 1960 to 1991, the sampling fraction was 25 per cent, that is, one in every four Brazilian households answered the extended questionnaire. In 2000 and 2010, the sampling fraction was reduced to 10 per cent. The survey redesign did not imply any loss of quality or reliability of the sample data. With the advent of the sample survey, the number of items covered for the entire population was reduced. Data from the population universe of the recent Census is only interesting when analysing highly disaggregated geographical levels. Thus, what researchers have for 2000 and 2010 are the micro-data from the 10 per cent sample and the aggregated data by census section, containing around 20,000 people each.

For the purposes of this research, only the micro-data from the samples of the Demographic Censuses will be used, due to the range of themes investigated in them and also because it is not our purpose to carry out spatial analyses with detailed intra-municipal (census sectors) data.

The time period of interest to us will be from 1960 to 2010. 1960² marks the beginning of the sampling and is also the first year for which we have micro-data available. By definition, the Censuses have nationwide coverage. The micro-data of the Censuses from 1960 to 2010 are available for free, public download from the *Centro de Estudos da Metrópole* (Centre for Metropolitan Studies – <http://www.fflch.usp.br/centrodametropole>).

Table 1
Overview of Available Datasets in Brazil

<i>Dataset</i>	<i>Frequency and comparability across years</i>	<i>Objectives</i>	<i>Key variables</i>	<i>Availability</i>
Population Census	Decennial: 1960 to 2010	Collect information on socio-demographic data about families and households	Gender; age; ethnicity; schooling; activity status; occupation status; income; migration.	Microdata
National Household Sample Survey (PNAD)	Inter-census years: 1967 to 1974; 1976 to 1979; 1981 to 1993; 1995 to 2011.	Collect socio-demographic information in years not covered by census	Census variables; new variables available in special supplements.	Microdata
Census of Agriculture	Decennial: 1920 to 1970 Quinquennial: 1975 to 1985; 1995/1996; and 2006.	Collect information on the economic activities performed by farmers and agricultural enterprises.	Characteristics of the agricultural producer and of the establishment; land use, agricultural production; employment and wages; revenues; implements; financing; debt.	Aggregated statistical tables
Annual Survey of Industry (PIA)	Yearly: 1966 to 1985; 1986 to 1995; 1996 to 2010.	Gather financial and economic information on industrial activities in the country and their evolution over time.	Employment; wages; costs and expenses; revenues; value added; investments.	Aggregated statistical tables
Annual Survey of Construction Industry (PAIC)	Yearly: 1990 to 1995; 1996 to 2010.	Gather information of the construction sector in the country	Employment; wages; investments; revenues; consumption by types of building materials; value of construction put in place.	Aggregated statistical tables
Annual Survey of Trade (PAC)	Yearly: 1988 to 1995; 1996 to 2010.	Gather information of the trade sector	Employment; wages; net sales revenues; trade categories; trade margin.	Aggregated statistical tables
Annual Survey of Services (PAS)	Yearly: 1998 to 2010.	Gather information of non-financial service activities	Employment; wages; revenues; operating expenses by service categories.	Aggregated statistical tables
Annual Report of Social Information (RAIS)	Yearly	Collect information on workers formally employed	Gender; schooling; occupation; type of employment; wage; hours of work.	Microdata

General Register of Employment and Unemployment (CAGED)	Monthly	Record the movement of the private employment	Private employees hired and fired.	Microdata
Urban Informal Economy (ECINF)	1997 and 2003	Collect information on informal activities in Brazilian economy	Employment; accounting; formalization; economic activity; financial results.	Aggregated statistical tables
System of National Accounts	Quarterly and Yearly: 1995 to 2012 (GDP since 1900)	Measuring the generation, distribution and use of income in the country	GDP; private consumption; gross fixed capital formation; government expenditure; imports and exports.	Aggregated statistical tables
Regional Accounts of Brazil	Yearly: 1995 to 2010	Estimation of regional GDP	GDP by geographic region and State; value added by activity.	Aggregated statistical tables
GDP of Municipalities	Yearly: 1999 to 2010	Estimation of municipal GDP	GDP by municipalities; value added by activity.	Aggregated statistical tables

Table 2

Overview of Available Datasets in India

<i>Dataset</i>	<i>Frequency and comparability across years</i>	<i>Objectives</i>	<i>Some key variables</i>	<i>Availability at this point</i>
Population Census of India	Decennial: 1881 to 2010	Collects data on key demographic, economic and social characteristics.	Name, birth place, sex, age, religion, mother tongue, literacy, place of last residence, marital status, fertility, economic status, principal and subsidiary means of livelihood, nature of industry, class of workers, distance travelled for work. Net area of land under cultivation/plantation and net area of irrigated land for cultivating households.	Micro data (Not all censuses contained all information)
National Sample Survey (NSS) -employment-unemployment survey	Thick rounds: 1983, 1987-88, 1993-94, 1999-2000, 2004-05, 2009-10, and 2011-12. Thin rounds: 2004, 2005-06, and 2007-08	Collects data on socio-demographic and employment-unemployment characteristics of household members.	Social group, religion of the household. Age, Gender, education, employment-unemployment status, earnings, industry of employment, occupation for household members. A brief consumption expenditure module of the household.	Micro data

National Sample Survey (NSS) – consumption survey	Thick rounds: 1983, 1987-88, 1993-94, 1999-2000, 2004-05, 2009-10, and 2011-12. Thin rounds: 1989-90 to 2007-08	Collects data on socio-demographic and detailed consumption expenditure of the household.	Social group, religion of the household. A detailed consumption expenditure module for the household.	Aggregated statistical tables from the NSS reports
National Sample surveys –other household surveys like All India debt and investment survey (AIDS) and situation of farmers survey (SFS)	1993 and 2003	To collect data on special aspects of households like debt, investment among households, and of cropping pattern and cost of operations among farmer households.	AIDS: value of land, livestock, building and other construction owned by the household, agriculture machinery, non-farm business equipment, transport equipment, durable assets, financial assets, cash loans and kind loans outstanding. SFS: costs, receipts, cropping patterns, access to modern technology, resource availability, and indebtedness of farmers and farmer households in India.	Aggregated statistical tables from the NSS reports
National Sample Survey – Informal Non-Agricultural Enterprises survey	1999-00	Gather information on informal non-agricultural enterprises	Type of enterprise (whether propriety, partnership, cooperative etc.), industry, whom they sell to, how/from whom they get their raw materials, value of fixed assets, cost of each input (including labour), operating expenses, outstanding loans, and value of receipt from sales.	Aggregated statistical tables from the NSS reports
Unorganized Manufacturing enterprise survey	2000-01 and 2005-06	Gather information on unorganized manufacturing enterprises	Same as above	Aggregated statistical tables from the NSS reports
Unorganized Services excluding Trade & Finance	2001-02 and 2006-07	Gather information on Unorganized Services excluding Trade & Finance	Same as above	Aggregated statistical tables from the NSS reports
Unincorporated non-agricultural Enterprises (manufacturing, trade and other services activities, excluding construction) survey	2011-12	Gather information on Unincorporated non-agricultural Enterprises	Same as above	Aggregated statistical tables from the NSS reports

Annual Survey of Industries	Yearly since 1960, except 1972.	Collects statistical on organized manufacturing sector comprising activities related to manufacturing processes, repair services, gas and water supply and cold storage.	Value of fixed, value of working capitals and loans, liabilities, receipts, expenses, inputs, products and by products. Man-days worked absenteeism, earnings and social security benefits separately for male and female workers, contract and regular workers, and production workers / supervisory and managerial staff.	Aggregated statistical tables from the ASI reports that are available in the website of labor bureau of India.
National council of Applied Economics Research surveys	1993-94 and 2004-05	To collect social, economic, fertility, education, income data from households in India	Income, employment, consumption expenditure; Health, education, fertility, family planning, marriage, and gender relations in the household and community; reading/writing and arithmetic knowledge tests for children; Height and weight measurement of children under age 5. Facilities assessment at primary school and primary health care facility. Assessment of employment opportunities and infrastructure in the village.	Micro data at the household level, school, health care facility, and villages.
National Accounts Statistics	National estimates: 1950-51 to 2011-12. State series from 1980-81 to 2011-12.		Gross National Income (GNI), Net National Income (NNI), Gross Domestic Product (GDP), Net Domestic Product (NDP) at factor cost by industry, and Per Capita Income (Per Capita Net National Income at factor cost (in rupees) for both current and constant prices.	Aggregated tables available in the website

Regarding the content of the questionnaires, it is important to highlight that there were large changes over the course of time. In the 1960 census, there were only 24 questions about individuals and 13 on the characteristics of their housing, while in the 2010 census there were 23 items on housing and 100 on the individuals. The changes did not consist only of adding questions, but even so a longitudinal comparison is still possible. There are several possibilities for standardizing and adapting the variables that we intend to adopt in this research, including some proposed by IPUMS (Integrated Public Use Micro-data Series, International) for censuses and surveys in several countries (including India). We will later jointly present the contents of the Censuses and PNADs.

Over time, some modifications were made to the work-related variables - Working Age Population, Economically Active Population, Occupied Workers, Unemployed and Workers by Status of Employment (according to ILO Definition) and Occupational Class Structure - although it is possible to maintain the inter-temporal comparison without greatly affecting the degree of reliability.

The main difference resides above all in the manner in which the variables “Condition of Activity” (economically and non-economically active population) and “Condition of Occupation” (occupied and non-occupied workers) are collected.³

Regarding the information on the individuals’ occupations, each Demographic Census adopted a different system for classifying and cataloguing the different types of job and activities. There are important lines of continuity from one census to the next, but they are not directly comparable with each other and not directly related to the international classifications. We suggest the use of the occupation and sector standardization made by IPUMS. The occupational variables of all the Censuses were converted to the ISCO-88 system (International Standard Classification of Occupation – 1988) at the one-digit level, which also allows a comparison with the Indian case (for which IPUMS carried out the same work). The same can be said for the sector of economic activity.

All the Censuses include the individuals’ income variables and, from 1970 onwards, the values were expressed in monetary units (not in category bands, as in the 1960 Census). From 1980, a difference started to be drawn between income from work and from other sources. Detailing and measuring these data continuously allows inequality to be measured and decompositions to be applied.

Throughout the whole period it is therefore possible to use the following labour and income variables, with a certain degree of reliability in the comparison:

- Working-Age Population: Persons aged 10 or over (1960-2010);
- Economically Active Population: bearing in mind the differences in data collection mentioned above (1960-2010);
- Condition of Occupation (occupied and non-occupied): – bearing in mind the differences in data collection mentioned above (1960-2010).
- Status of employment (Employers; Employees; Self-employed; Unpaid Workers);
- Formalization (wage-earners): differentiation of types of contracts (public, private, etc., from 1980-2010);

- Occupational Structure(1960-2010)
- Sector of Economic Activity (1960-2010);
- Average Individual income (1960-2010);
- Average labour income (1980-2010);
- Total and *per capita* household income (1970-2010).

(ii) The importance of PNAD 1976

The 1976 National Household Sample Survey (PNAD – discussed in more detail in section 3) will be taken to represent the labour market and the occupational structure at the end of the 1968-1973 expansion cycle of the Brazilian economy (Singer, 1981). This allows us to move forward in the investigations, as it contains a much greater number of questions than the Census. Besides the variables listed above and contained in the Census, the PNADs also include questions on topics such as:

- Access to work;
- First employment;
- Time out of work (both for those who sought and for those who did not seek work);
- Income in kind;
- Social Security contributions;
- Duration of Job Tenure;
- Activities carried out by those out of work;
- Last occupation (occupational mobility);
- Job seeking (time and means by which work was sought);
- Sources of income unrelated to work (rent, retirement pensions, other pensions, donations, others).

(iii) Using the Available Data

The availability of census micro-data allows us to estimate inequality measures (Gini coefficient, Theil index, mean log deviation, coefficient of variation, etc.), and to use different quantitative methods to explore inequality (ANOVA, decomposition techniques, etc.).

Moreover, the inequality measures can be decomposed by other variables, such as sex, educational level, race (with the exception of the 1970 Census, when the race/colour question was not included in the questionnaire), geographical regions and units of the Brazilian Federation. In spite of great transformations in the Brazilian education system throughout the period, the variables pertaining to the educational level, can be dealt with in terms of years of schooling (continuously or in bands),⁴ or on the basis of transitions between different stages of the education system (categorical or discrete variable).

2. India

(i) Census of India

Localized censuses have been conducted in India since the early 1800s, but the first all-India census was conducted in 1881. Since then there has been a decennial census and the

last one, the 15th Indian census, was conducted in 2011. The census of India collects data on every individual in every household, including information such as name, birth place, sex, age, religion, mother tongue, literacy, place of last residence, marital status, fertility, economic status, principal and subsidiary means of livelihood, nature of industry, class of workers, distance travelled for work and much more.

Not all variables are available from 1881 through 2011. New questions were introduced in more recent censuses and some questions that were in earlier censuses were dropped. For example, in the 2001 census the following are some of the new questions that were included:

- Information regarding age at marriage (which was till then collected only for females) was collected for males.
- A question on total or partial disability was included (in the census of 1981 the question on disability was canvassed but it was dropped in 1991).
- Those engaged in non-agricultural activities were also asked how far they traveled to work.
- Net area of land under cultivation/plantation and net area of irrigated land were also collected for those households who were engaged in agriculture.

Achieving comparability between censuses in terms of concepts of work poses some difficulties, though several scholars, such as Daniel and Alice Thorner and J Krishnamurthy as well as the 1961 Census officials, have made it possible to use the data to make comparisons over time, subject always to some caveats. The size and structure of the working force has been analyzed for the period 1881 to 1961, even though figures for some intermediate years were problematic.⁵ The trends and patterns can be analyzed for males and females (with less confidence) for broad branches of economic activity as well as in further detail. Agricultural workers are divided between cultivators and agricultural labourers and those engaged in livestock, forestry and fishing etc..

The 1961 Census is sometimes regarded as the best in the post-Independence period. Subsequent censuses continue to have some problems of consistency in the definition of work, with the measured female labour force participation rate, for example, changing from census to census in implausible ways.

Summary data at country, state and district levels of some key variables for the 2001 and 2011 census are available in the tabulated form in the website of the census of India (<http://www.censusindia.gov.in/>). Tabulated data for older censuses can be obtained or purchased at a very low price from the Census office in New Delhi. Individual record level data can be purchased from the Office of the Registrar General and Census Commissioner, Ministry of Home Affairs, Government of India. Micro individual level data from the 2001 Census is available at a workstation at Jawaharlal Nehru University (the name of the individual and other sensitive fields are anonymized).

(ii) Other Historical Data

Apart from the censuses, considerable historical data exist on variables relevant to the analysis of long term patterns of inequality in India, but they are subject to several problems, especially for the pre-Independence period.

- No official estimates of inequality exist for the period before 1950. Nor are there unofficial estimates specifically of inequality of assets or income or consumption.
- Series data (i.e. comparable over time) on such variables as consumption, investment and national income do not exist or have had to be constructed later from scrappy data.
- Data are of variable quality and coverage
- It is not possible today to scrutinize the raw data and assess their quality. We do not always know whether they are reflective of the true situation or are biased to reflect the preferences of the colonial rulers.
- The data that are available are already processed and it is not possible for the present project to go back and build new estimates.

There are thus some serious problems in comparing the much richer data sources of more recent years with the data for the pre-Independence period.

Before Independence there are a variety of sources on prices and wages, although there are problems of consistency and of bias in government statistics. After Independence, a major survey was the Agricultural Labour Enquiry in 1950-51. This was repeated in 1955-56 and followed by periodic Rural Labour Enquiries starting from 1963-64. There have so far been eight Enquiries, the latest being in 2009-10. While there were some problems of changes in concept and coverage in the earlier Enquiries, these provide a good source for examining changes in the earnings of rural and agricultural labour. After Independence the National Sample Survey also started to provide information on wages and employment at the national level, starting in 1955-56 (9th round).

Given that land is the major factor of production in rural India, patterns of landownership and operation are an important indicator of inequality levels and changes over time. A pioneering analysis of inequality in landownership was done by Dharma Kumar for the Madras Presidency covering the period 1853 to 1947. After Independence several surveys have attempted to examine the distribution of land by ownership and operation. The first sample survey was in 1953-54 by the National Sample Survey (NSS) and there have been periodic surveys until the present time. An agricultural census has been conducted since 1970-71 on a five-yearly basis but the data are not regarded as being very reliable by many analysts. The NSS source makes it possible to look at the trend since Independence in landholding distribution. The series extends to at least 2002-03. Using these data, trends in the ownership and distribution of land over time can be examined in some detail to throw light on whether land is becoming more unequally distributed in terms of ownership or operation.

Several attempts were made before Independence to measure poverty among the working population. These provide some idea of the living standard of the poor but do not give a measure of the living standard of the non-poor, and hence do not permit proper measurement of inequality of income. Nutritional surveys conducted in the 1930s and early 1940s indicate that 30 per cent of the population got less than the nutritional norm of 2300 calories per consumption unit. However the surveys covered only about 1500 families and cannot be generalized with confidence. After Independence poverty estimates are available using the consumption data collected by the NSS from 1950-51 onwards; the first official poverty

estimates date from 1960-61.

3. Overall comparability

The censuses in the two countries are held at similar intervals and some of the information is common, notably on economic activity and occupation. They are clearly basic sources for long term analysis in both countries, and many variables can be compared directly. Nevertheless there are important differences, in that the Census data are much richer in Brazil, especially the sample.

In particular, data on economic activity are much more restricted in India. In the Brazilian data, not only is there more complete information on occupation and work status, there are also income data. So a variety of measures of inequality can be obtained. The Indian Census, on the other hand, provides very little information on the extent of inequality (with the possible exception of landholding for cultivating households).

Another advantage of Brazilian Census data is that it provides information on race (or at least self-defined skin colour). Caste inequality in India is in some ways parallel to racial inequality in Brazil, but the Indian Census has not collected general information on caste since 1931. Only the relatively deprived categories of Scheduled Caste and Scheduled Tribe are identified. The recent caste census will not provide useful data in time for this project.

So the Indian Census cannot provide as good a long term historical picture as the Brazilian Census, and alternative data sources are required in the Indian case. Some of the more important alternative sources are described above, especially the Rural Labour Enquiries and the early rounds of the National Sample Survey, but they will require assessment on a case by case basis.

The use of PNAD 1976 to provide a baseline for the economic structure at the end of the cycle of expansion of the Brazilian economy would not have an exact counterpart in India. Either the 1983 or the 1987-88 NSS could play a similar role since this was before the impact of economic reform at the end of the phase of the “Hindu rate of growth”.

III. REGULAR NATIONAL LEVEL SURVEYS ON LABOUR MARKETS

1. Brazil

(i) The Design of the National Household Sample Surveys (PNAD)

The National Household Sample Surveys (PNADs) are currently carried out on an annual basis by the IBGE. They are the main surveys on work and income in the country.⁶

The PNAD started in 1967, covering only the Northeast, Southeast, South and the Federal District (leaving out the Centre-West and Northern regions). Until 1969, it was carried out on a quarterly basis. In 1970, it was interrupted because of the Census, and since then, this has not happened in all Census years – that is, the PNAD is not implemented in Census years. Otherwise, from 1971 onwards, it was carried out annually. In 1973, its geographical scope widened to include the urban areas of the Centre-West and Northern regions. This design remained until the end of the 1970s. In 1981, there was the inclusion of the Centre-West rural zones, a sample design that remained until 2003. Lastly, in 2004,

the rural zones of the states of the North (Acre, Amazonas, Pará, Rondônia, Roraima and Amapá) were included. From that year onwards, the PNAD has become representative of the whole of Brazil.

At a more disaggregated level, the representativeness of the PNADs includes the urban and rural areas of the units of the Brazilian Federation and the nine metropolitan regions: São Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Curitiba, Salvador, Recife, Fortaleza and Belém. The survey design is based on a sample in several stages. In the first stage, the country is roughly divided into strata. Within the strata, there is a random selection of clusters with probability proportional to the population size. In the second stage, census sectors are drawn at random and, lastly, in a third stage, the households are selected. The households are always drawn by means of systematic sampling with equal probability. In order to permit estimates to be made for the population as a whole, weights are calculated for both households and individuals.

Nascimento Silva, Pessoa and Lila (2002), responsible for the IBGE's research methodology, have developed a methodology for incorporating the sample design of the PNADs in the calculation of measures of dispersion by identifying the strata and clusters of which the households and individuals are a part. The adjustment procedure indicated by the authors may be implemented through general-use statistical packages (R, Stata, SPSS).

Regarding the contents of the survey, the PNAD questionnaires are divided into two parts: basic and supplementary questionnaire. The structure of the basic section essentially remains constant during each decade, whereas the questionnaires with supplementary themes vary year by year. We will not be dealing with the themes of the supplementary questionnaires in the present research. The basic questionnaire contains items on demographic characteristics of the population, education, labour market, migration, child labour and fertility.

The PNADs have much more extensive questionnaires than the Demographic Censuses. The sections on the labour market alone in the PNAD are larger than the any round of the Census.

(ii) The content of the PNADs and their relationship to the Census data

Because they are produced by the same institution, the questionnaires of the Censuses and PNADs have some similarities. As the PNADs are aimed at producing information in the periods between Censuses, there is a high degree of comparability throughout the years.

Regarding the characteristics of the households, it is possible to identify consumer items (possession of goods and access to services), housing conditions and conditions of the surrounding areas (Cavenaghi, 2010). There is a basic identification section in which information on the individual's demographic profile is collected. The question on race/colour appears in almost all of the Censuses, with the exception of 1970, whereas it only became regular in the PNADs from 1987, although it was present in some previous rounds as part of the supplementary questionnaire (Osório, 2003). In addition, information on family structure (nuclear, single parent, with or without children, etc.) has been also collected. In the 2010 Census, the possibility of more

than one individual being responsible for the household was included. Since 1991, the Censuses have included questions on disability and special needs and from 2000 onwards, information on this topic has been collected in a more detailed fashion. The 2010 census also included a section on indigenous populations. This set of questions is answered by those who declare themselves indigenous in the question about race and for all of those who live in territories defined as indigenous.

The treatment of education remained quite stable in the Censuses and PNADs (Rigotti, 2004), and it is possible to keep track of schooling levels and school attendance. Only in the 2010 Census can the calculation of years of schooling be made by means of a discrete quantitative variable, but categorical variables can be constructed on levels of schooling that are comparable for all the periods.

Looking at the questions on demography, we can observe that the topic of fertility was covered in all the Demographic Censuses and, from the 1990s; it was systematically repeated in all PNAD rounds (as was also the case for migration and child labour). The questions on nuptiality were present in all Censuses and in a good part of the PNADs until the early 2000s, when they were removed from the questionnaire. The migration topic was included in all Censuses from 1960 to 2010 and in several PNADs. Mortality was not investigated by the PNAD and only appears in the 1980 and 2010 Census questionnaires. Researchers who deal with this theme would generally rather use the corrected statistics from the Civil Register.

Regarding the topic of labour, the older Censuses investigated only individuals' "habitual work". More recently, this concept was replaced by the notion of work undertaken in the week prior to the survey date. Aside from the issue of the reference period, the comparability of the labour data depends on the occupational classification system and on the sectors of economic activity. Each Census employed a different system with increasing detail and complexity, and the PNADs always used the systems employed in the Demographic Censuses at the beginning of the decade. Numerous statistical tricks can be used to make the classifications compatible. As we have already mentioned, in this research we will make use of the IPUMS proposal for mapping the occupational measures onto the ISCO-88 system (International Standard Classification of Occupation – 1988) and the industrial sector onto ISIC-3 (International Standard Industry Classification), both produced by the International Labour Organization. The same standardization is available for the Indian data (In India, the mapping has already taken place. That is, the NIC is already coded based on the ISIC and we need not manually do it. See page 9 in this document - http://mospi.nic.in/mospi_new/upload/nic_2008_17apr09.pdf).

Regarding income, given the changes in the monetary unit and the inflation present throughout the period, we will use the deflators and converters suggested by Corseuil and Foguel (2002). The procedure indicated in these authors' studies consists in a method for enabling intra-national monetary comparisons and can be extended to the other years, up until 2010. We have yet to establish the conversion standards for international comparison.

As in the case of the Census, the availability of micro-data allows the use of several methods in the study of income inequality and its evolution over time, particularly in years

not covered by Census.

Thus, taking into account the observations described above, the variables listed at the beginning of this text allow both databases (Censuses and PNADs) to be treated in complementary fashion throughout the period studied. The present study will use the 1960, 1970 and 1980 Censuses to present the main transformations in question throughout the earlier period under study, with the 1976 PNAD as a point of reference for the end of the expansion cycle of the Brazilian economy, as previously mentioned. The PNADs from that year onwards will be used – including those of 1981, 1989, 1993 and 1995-2011 – for the analysis of more recent trends.

2. India

(i) The National Sample Survey Organization (NSSO) surveys

The National Sample Surveys are conducted by the National Sample Survey Organization (NSSO), Ministry of Statistics and Programme Implementation, Government of India. The NSSO was set up in 1950 to obtain information on key socio-economic and demographic variables in India. These are annual cross-section surveys conducted in rounds, and each round is usually for one full year. These Socio-Economic (SE) surveys cover a number of topics, the most important of them being consumer expenditure and employment-unemployment. As noted above, the first employment-unemployment survey was carried out in 1955-56; the widely used quinquennial surveys of employment-unemployment and consumer expenditure were started in 1972-73. There are also less frequent but more specific surveys on topics such as migration, indebtedness, and informal enterprises.

The SE surveys are conducted on a ten year cycle. In this cycle, the subject of land and livestock holdings, debt and investment is allotted one year; social consumption (for example on education and health care) is allotted one year; quinquennial surveys on household consumer expenditure, employment & unemployment are allotted two years; and four years are allotted to non-agricultural enterprises, namely, manufacturing, trade and services in un-organized sector. In the other two years subjects of current/special interest on the demand of Central Ministries, State Governments and research organizations are covered. The NSSO surveys are considered the most representative all India surveys. The whole of the Indian Union is covered except for interior villages of Nagaland and inaccessible villages in Andaman & Nicobar Islands.

(ii) Employment-unemployment and consumption surveys

Quinquennial rounds of the NSSO surveys are conducted on the subjects of employment, unemployment and consumption. These are “thick” surveys wherein more than one hundred thousand households are sampled. Thick rounds of both employment-unemployment surveys and consumption surveys were conducted in 1983 (38th round), 1987-88 (43rd), 1993-94 (50th), 1999-2000 (55th), 2004-05 (61st), 2009-10 (66th), and 2011-12 (68th round – an additional round was devoted to employment data in 2011-12 outside the quinquennial cycle). NSSO also conducts thin employment surveys on about 50,000 households or less. Thin rounds on employment-

unemployment were conducted in 2004, 2005-06, and 2007-08 and consumption data were collected annually from 1989-90 to 2007-08, except in the years that have a thick round.

The consumption surveys collect data on individual demographics and education of household members, and household level detailed consumer expenditure. The employment-unemployment survey collects data on individual demographics and education of household members, and employment details such as industry, occupation, type of work (casual/regular), characteristics of workplace and job contract, earnings, and length of unemployment, for all working members. At the household level details such as religion, social group, and a less detailed consumer expenditure module are available. The definitions of important variables (such as wages, employment, and consumption) are similar across different rounds and are comparable.

The design of the socio economic survey was revised in the 27th round (1973-74) and the methodology adopted in that round was used in subsequent rounds, albeit with important changes in stratification and sub-stratification (National Sample Survey Organization, 2004). The basic design however remained the same. States are divided into rural and urban sectors. Within the rural sector and urban sector, strata and sub strata are identified. Census villages in the rural sector and Urban Frame Survey (UFS) blocks (mainly) in the urban sector are the first-stage units. Households are the second-stage units. Taking into account changes in the Indian population and urbanization, the rules for stratification and sub-stratification in both rural and urban sector changed in most subsequent rounds. To measure consumer expenditure more accurately, in the 43rd round a separate stratum was formed comprising areas where the affluent households of big cities lived. There were also changes in the method of sampling first stage units and households in many rounds. Despite these changes across different rounds, estimates at the national and state level are easily obtained because the sample weight in each round for every household is calculated and provided by the NSSO along with the raw data. For the thin rounds, the selection of first stage units was based on the major subject of inquiry that year.

All thick rounds and thin employment-unemployment rounds are available with IHD up to the 2011-12 round. Consumption rounds are not available with IHD but it should be noted that a brief consumption module is included in the employment-unemployment survey as well. Hence, to the extent that earnings are the main focus of our analysis, consumption data can be used from the employment survey and analyzing the consumption survey separately will not be necessary.⁷

Many interesting analyses can be done with the NSS data. Inequality of earnings and working conditions can be measured using the employment surveys. One could also analyse inequalities arising from labour market access as determined by caste (SC and ST only) and social group. Some studies using the NSS data include Das (2012), Sarkar and Mehta (2010), and Azam and Prakash (2010). Consumption inequality can be measured using the consumption surveys or from the consumption module in the employment surveys.

The main advantage of the NSS data is that they are nationally representative. Further, the thick NSS consumption and employment surveys are the only data on households that are

representative at the state and regional levels. Obtaining estimates below the state level is tricky using NSS.⁷ NSS analysis is best not done at the district level, because the boundary of the districts has been changing and the number of districts has been increasing. This creates difficulty in comparing districts over a period of time. A good mapping of changes in district boundaries is required to carry out an empirical analysis. Moreover it is not clear whether we can get reliable estimates at the district level because of the small sample sizes at such a disaggregated level. However, NSS data are representative (and statistics can be estimated) at the regional level. Regions are groups of districts at the sub-state level. To the extent that NSS regions are a good geographical area to study inequality, we can conduct regional analysis. But this kind of analysis also merits a closer review because NSS regions do not separate urban and rural areas, which is problematic because the urban landscape is very different from the rural. NSS data are not subject to seasonal differences because every state is surveyed all across the year.

However, the NSS surveys may be unreliable for various reasons. (1) Some definitions and measurements are not fully consistent in every survey, making comparisons over time and an empirical time series analysis difficult. For example, there seems to be inconsistency in the measurement of labour force participation of women in different rounds of the employment surveys. Further, the 1999-2000 wage data has “discrepancies” (shows higher wage growth rate than other years); (2) some important variables are affected by erratic fluctuations, both climatic and economic. For example, if the survey was conducted in a bad agricultural year or the country in general is facing a macroeconomic shock, data from those years may not be representative of the longer term situation. For example, 1987-88 was a bad agricultural year and the estimates from that year are also not reliable. It might be a good idea to use thin rounds to supplement thick rounds in such cases.

Because the NSS surveys of employment and unemployment were (at least until 2011-12) only carried out every 5 years, the Labour Bureau of the Ministry of Labour has recently launched an annual survey of employment and unemployment. However seems to add little compared with the NSS, and since it is new cannot be used to analyse changes over time. It is therefore not used in our study.

3. Overall Comparability

The PNAD and the NSSO employment and consumption surveys will form the backbone of the empirical analysis under our study, since both provide nationally representative estimates over long periods of time. They have a number of common variables, including details of wages and employment, unemployment and workplace characteristics. While there are differences in occupational and industrial classifications, adjustments for approximate comparability are possible. Perhaps more problematic, the measures of labour status – the type of labour contract, informality, etc. - differ, in part because the labour market realities also differ. Thus there is no counterpart in India to the presence of a signed labour card in Brazil. So it will often be difficult to undertake identical analyses in the two countries – and even if identical variables are

used, the interpretation must take into account the very different situations.

One important difference concerns the lack of availability of income data in the NSS surveys, while this is available from the PNAD. Wage data are available in the NSS, but it is essential to also include some estimate of income from self employment, since half of the working population is self-employed. On the other hand, the NSS includes consumer expenditure data, which is not available in the PNAD. This poses significant difficulties for comparative analysis, which will have to be overcome by using additional data sources, for instance to estimate the relationship between income and consumption and apply this as a proxy.

The lack of annual NSS data is less of a problem, since long term trends can in principle be captured by quinquennial surveys, so that in the case of the PNAD it may be sufficient to take one survey year in five as well.

IV. OTHER SURVEYS AND CENSUSES

This section presents a selection of other data sources likely to assist in analysing labour market inequality, including both enterprise and household surveys. Most of these provide some complementary information not available from the Census, the PNAD or the NSS surveys.

1. Brazil

Most of the data sources listed here are enterprise surveys or censuses carried out by IBGE.

(i) Census of Agriculture

The Census of Agriculture collects information on the agricultural and livestock sector in Brazil.⁸ The census began in 1920 and was repeated every 10 years until 1970 and every 5 years after that. There was no census in 1990 or 2000 and the last two censuses available are for 1995/1996 and 2006.⁹

The census includes all agricultural and livestock establishments in Brazil, regardless of size, legal constitution and location, and collects information on the person responsible for the establishment/producer (age, sex, schooling, legal status of the producer and the land, etc.), on the establishment itself (personnel employed; area; use of land, irrigation methods, silos and warehouses; use of mechanical/animal power; use of agrochemicals, use of organic agriculture, etc.) and on the production (main crops; milk production; egg production; private/government funding; revenue; debts; investments; value of the goods; etc.).

Furthermore, the Census of Agriculture makes available variables computed from information captured directly, such as, for example, the Gini index (by municipality and unit of the Brazilian Federation), which allows one to track the concentration of land use between 1985 and 2006.

Regarding the gathering of information on occupation and wages, the Census provides a more disaggregated differentiation of the type of agrarian worker than the PNAD, especially

in 2006. Thus, it is possible to identify the number of workers who have family ties with the producer/person responsible, as well as if the employee is permanent or temporary or in any other employment relationship. The Census also covers workers who do not participate in any agricultural activities in the establishment, as well as the domestic workers.

(ii) Annual Industrial Survey

The Annual Industrial Survey (PIA), which is implemented by the IBGE, has collected information on Brazil's industry since 1966. Initially a census, from 1981 it adopted a process of probabilistic sampling, which reduced costs, and allowed for greater ease when collecting information and greater speed when publishing results. In 1996, the PIA was reformulated and disaggregated into two surveys: Annual Industrial Survey – Companies (PIA-Empresa) and Annual Industrial Survey – Production (PIA-Produto). PIA-Empresa is the survey that corresponds to the former PIA and collects the economic and financial characteristics of industrial companies, while PIA-Produto focuses on industrial services and products.

The PIA sample can be currently disaggregated into two groups:¹⁰ *i*) direct selection of all the companies with 30 or more employees and/or revenue over a certain annually stipulated level – in 2010, it was R\$ 9.33 million (approximately US\$ 4.7 million); and *ii*) probabilistic sampling of companies with fewer than 30 employees. In 2010 the PIA had a total of 52,814 companies in its sample, of which 37,808 belonged to the first group.

The IBGE applies different questionnaires to the two sample groups; the first answers a complete questionnaire and the second a simplified questionnaire.¹¹ They both include information such as:

- Employed personnel (employed at the end of the year of reference and average number of employees in the year): disaggregated into non-salaried personnel; salaried and linked to production; and salaried and not linked to production.
- Costs and expenses: expenditures on personnel (salaries, salary advance and other kinds of earnings); purchases of raw material; direct production costs; stocks.
- Other costs and expenses: rent; depreciation and amortization; other operational costs (transportation, advertising; office material, etc.).
- Net sales revenue: disaggregated into gross revenue and deductions (cancelled sales, taxes, etc.).
- Other revenue: financial revenue; acquisitions, changes in tangible assets made in the year: purchase and improvement of assets, including adjustments and deductions of depreciation accounts.
- Variables derived from information collected from the company: manufacturing value; intermediate consumption; value added; profit; etc.

For the PIA-Produto, the main variables included are:

- Net sales revenue from industrial products and services produced by the local unit.
- Information on the products made and industrial services provided by the local unit: code, description and unit of measurement of the product; quantity produced per year; quantity sold per year.

The PIA information can be decomposed by field of industrial activity, according to the National Classification of Economic Activities (CNAE), ‘number of employees’ band and geographical macro region. The regional breakdown by unit of the Brazilian Federation is limited to the branch of activity – CNAE two digits. Lastly, for companies with 30 employees or more, the information can be disaggregated by CNAE (four digits) and by municipality.

(iii) Annual Survey of Construction Industry

The Annual Survey of Construction (*Pesquisa Anual da Indústria da Construção – PAIC*) is another IBGE survey. Between 1990 and 1995, the PAIC included all of the construction companies whose production in this activity was at least 80 per cent of the total. From 1996 to 2001, the sample used consisted of all the companies in the sector with 40 or more employees. Lastly, from 2002, the PAIC adopted a process of probabilistic sampling similar to that of other IBGE sector surveys.

The PAIC sample can be disaggregated into two groups along the same lines as the PIA,¹² so that out of a total of 138,000 construction companies in Brazil in 2010, the PAIC sample included 19,931 companies, of which 12,159 belong to the first group.

The information collected by the PAIC is generally similar to that in the PIA, with the exception of some items specific to the construction sector. Thus the main PAIC variables can be summarized as: employed personnel and salaries; gross revenue; costs and expenses; asset acquisition and downtime/write-off; construction materials consumed; value added; value of building works and/or services; types of building works and/or construction services per customer, etc.

The PAIC information can be disaggregated by ‘number of employees’ band, geographical macro region and industrial activity branch.

(iv) Annual Survey of Trade

The Annual Survey of Trade (*Pesquisa Anual de Comércio – PAC*) has been carried out by the IBGE since 1998 with the aim of collecting information on the wholesale and retail sector in Brazil.¹³ Initially, the PAC used probabilistic sampling for approximately 60,000 companies, but had to reduce this figure between 1992 and 1995 due to budget restrictions, so small and micro companies were excluded from the sample. Only from 1996 did the PAC return to its initial aim of being representative of the universe of companies in the commercial sector.

Currently, the PAC sample can be disaggregated into three main groups:¹⁴ *i*) a census of companies with 20 or more employees; *ii*) a census of companies with fewer than 20 employees, which operate in more than one state of the Brazilian Federation, or which have an average revenue similar to those with more than 20 employees; and *iii*) a random sample of companies with fewer than 20 employees.

In 2010, the PAC covered the whole of Brazil,¹⁵ involving approximately 70,000 companies in the sample, of which 50,000 belonged to the first group, representing a universe of 2.031 million companies of Brazil’s sector. In addition, the IBGE applied different questionnaires according to the size of the company – a complete one for those with 20 employees or over and a simplified one for those with fewer than 20 employees.

The set of variables investigated by the PAC includes, among others: employees; personnel expenditures, including salaries, withdrawals and other earnings; net resale revenue; sales destinations; purchases; cost of merchandise; stocks; financial expenses; acquisitions and sales of tangible assets; trade/sales margin; internet sales system; telesales; door to door etc.

The PAC permits a breakdown at the CNAE two digit level (three digits for some activities), by company size, geographical region and by unit of the Brazilian Federation of the company's headquarters.

(v) *Annual Survey of Services*

The Annual Survey of Services (*Pesquisa Anual de Serviços* – PAS) has been carried out by the IBGE since 1998 with the aim of collecting information from companies of the non-financial services sector in Brazil. The PAS sample is similar to that of PAC, that is, a census of companies with 20 employees or more and probabilistic sampling for others.¹⁶ In 2010, the PAS covered approximately 90,000 companies,¹⁷ of which 50,000 belonged to the first group and were representative of a universe of 1.3 million companies from the Brazilian services sector. As in other cases, the IBGE applies different questionnaires according to company size.

The set of variables investigated by the PAS is similar to that of PAC, including, among others: employees, personnel costs, salaries, withdrawals and other remunerations, net revenue from services provided; destination of sales; spending on materials; financial expenses; acquisitions and write-offs of tangible assets.

Lastly, PAS permits a breakdown at the CNAE two digit level (three digits for some activities), by company size, geographical region and unit of the Brazilian Federation of the company's headquarters.

(vi) *Urban Informal Economy*

The Urban Informal Economy (ECINF) is an IBGE survey that aims to collect information on the informal activities in the Brazilian economy. ECINF was conducted twice – 1997 and 2003 – and implemented through a probabilistic sample of the target population: occupied households living from informal businesses.

The informal sector covered by ECINF includes all the economic units owned by own-account workers and employers with up to 5 employees, living in urban areas. Therefore the survey did not cover non-agricultural activities in rural areas or illegal activities because those activities cannot be easily recorded.

The concept of informality in ECINF follows the recommendations of the 15th International Conference of Labour Statisticians (1993) that considered the boundaries of informal sector in terms of the economic activity of the production unit and not the individual worker or the activity he or she undertakes.

The results of ECINF have been released for Brazil, geographical regions, federative units and metropolitan areas. They include information about the characteristics and financial aspects of enterprises in the informal sector, such as workers, revenues, costs, average profit, accounting system, legalization; and characteristics of workers (self-employed, employers,

and wage-earners), such as sex, age, schooling and wage.

(vii) *Use and availability of these data sets*

The ECINF, Census of Agriculture and sector surveys offer a huge variety of information regarding the structure of companies and production/revenue generated, which can serve as a basis for drawing up indicators for analyses and projections.

However, IBGE does not make the micro-data of the Agricultural Census and its sector surveys available, but only the aggregated information of the variables surveyed. Therefore, with the exception of the tabulations already in existence, it is not possible to make other cross tabulations of variables. Nevertheless, existing tabulations provide some relevant information for the study of inequality, such as the relationship between wages and earnings of employees and production, among others.

2. India

(i) *Annual Survey of Industries*

The Annual Survey of Industries is conducted by the Ministry of Statistics and Programme Implementation, Government of India and is the principal source of statistics on industries in the organized sector. These are factory or establishment level data. The ASI contains a census component and a sample component. The Census component covers all industrial units having 100 or more workers and all factories covered under “joint returns”. In less industrialized states like Manipur, Meghalaya, Nagaland, Sikkim, Tripura, and Arunachal Pradesh, all industrial units are included. A random sample is taken of those factories that are not in the census component. The reference period for the data is usually a year, running from April 1 to March 31.

ASI covers factories registered under Sections 2m (i) and 2m (ii) of the Factories Act, 1948, i.e., only factories having more than 10 employees with electricity or 20 employees (with or without electricity). Bidi and cigar manufacturing establishments registered under the Bidi & Cigar Workers (Conditions of Employment) Act, 1966, and all electricity undertakings engaged in generation, transmission and distribution of electricity, registered with the Central Electricity Authority (CEA), were covered under ASI irrespective of their employment size. Certain service units and activities like water supply, cold storage, repairing of motor vehicles and other consumer durables like watches etc. are covered under the survey. Service sector industries like motion picture production, personal services like laundry services, job dyeing, etc. are covered under the survey but data are not tabulated, as these industries do not fall under the scope of the industrial sector defined by the United Nations. Excluded are defence establishments, oil storage and distribution depots, restaurants, hotels, cafés and computer services and the technical training institutes, etc. (Source: verbatim from http://mospi.nic.in/mospi_new/upload/asi/all_about_asi.htm).

The list of registered factories/units maintained by the Chief Inspector of Factories (CIF) in each state and authorities in respect of bidi and cigar establishments and electricity

undertakings form the basis for the sampling frame of the ASI. The frame is regularly updated by the regional Offices of the NSSO in consultation with the Chief Inspector of Factories in the state.

ASI data are provided by the government in four formats. First, the detailed unit level data in which all the details as provided in the ASI questionnaires submitted by the factories are made available. Detailed unit level data are available for the years 1983-84, 1983-84, 1989-90, and 1993-94 to 2008-09 except 1995-96. Second, the summary unit level data - for earlier years all the information obtained in the ASI schedules were not processed but only the summary characteristics were tabulated. Data are still provided at the factory level but all details with a particular module are not made available. Only aggregates are available. This summary unit data contain the aggregate values of about 74 characteristics but the values of their constituents are not available. Summary unit level data are available for the years 1974-75 to 1994-95 except 1975-76 & 1978-79). Third, panel data at the factory level are made available for the years 1997-98 to 2007-08. Unit data, summary unit level data, and panel data are all sold by the Government of India for researchers and institutions. Fourth, tabulated data based on unit level data are available in the ASI reports and made freely available but more disaggregated tables - disaggregation being at the industry level and State levels, can be purchased if needed (http://mospi.nic.in/mospi_new/upload/asi/ASI_main.htm?status=1&menu_id=88).

ASI questionnaires contain detailed information on establishments. Part I of the questionnaire contains information on fixed assets (such as gross and net value of land, buildings, plant & machinery, transport equipment, computer equipment including software etc. used in the factory); value of working capital and loans, liabilities, receipts, expenses, inputs, products and by-products. Part II contains data on man-days worked, absenteeism, earnings and social security benefits separately for male and female workers, contract and regular workers, and production workers/supervisory and managerial staff.

Using ASI data, the functional distribution of income can be analyzed at the enterprise level. For instance, we can study the wage and profit shares and how they vary over time and across industries. Using unit data (and more effectively panel data), inequality within the firm or industry can be analyzed. For instance, wage gaps across male/female; production workers/managerial staff; contract/non-contract workers etc. can be studied over time and across firms and industries. Cause and effect phenomenon can also be studied. An example is to study the effect of technological change on inequality - the usage of computers may decrease the demand for monotonous tasks that can be automated, and depress the wages for the unskilled. This will increase the unskilled-skilled worker wage gap.

The disadvantage of ASI is that since only registered factories and some other establishments are covered, in practice only a small percentage of workers can be captured from this data source. Nevertheless, ASI is the most important source for understanding inequality in the formal sector.

(ii) *NSS survey rounds on topics other than employment and consumer expenditure*

Enterprise surveys: NSSO has conducted a number of enterprise surveys. In recent years these include Informal Non-Agricultural Enterprises survey (1999-00), Unorganized Manufacturing¹⁸ enterprise survey (2000-01 and 2005-06), Unorganized Services¹⁹ excluding Trade & Finance (2001-02 and 2006-07), and the unincorporated Non-agricultural Enterprises (manufacturing, trade and other services activities, excluding construction) survey (2011-12). There are differences between these surveys in terms of the subject matter depending on the industry, but all of them collect some basic data on small and medium enterprises such as, type of enterprise (whether propriety, partnership, cooperative etc.), industry, whom they sell to, how/from whom they get their raw materials, value of fixed assets, cost of each input (including labour), operating expenses, outstanding loans, and value of receipt from sales.

Other socio-economic household surveys: All-India household debt and investment surveys were conducted in 1992 and 2003. They collected information on value of land owned, livestock, building and other construction owned by the household, agricultural machinery, non-farm business equipment, transport equipment, durable assets, financial assets, cash loans and kind loans outstanding. The situation assessment survey of farmers conducted in the same years collect data on costs, receipts, cropping patterns, access to modern technology, resource availability, and indebtedness of farmers and farmer households in India.

These surveys permit the measurement of several aspects of inequality. Household asset inequality can be measured using the all India Debt and Investment surveys. Wage shares and functional distribution of income can be computed using the informal enterprise surveys. This needs to be supplemented with the same analysis for registered manufacturing establishments using the ASI data (see above). The enterprise surveys are a good source of information about the self-employed. The enterprise survey on the services sector seems to be the only detailed survey that asks information on all aspects of a service enterprise. This will be very useful in decrypting the black box of the service sector.

(iii) *National Council of Applied Economic Research (NCAER) Surveys*

The NCAER is a leading think tank which has conducted a number of nationwide surveys. Among these, the Human Development Profile of India (HDPI) and the India Human Development Survey are the most useful for analysis of income and other inequalities. NCAER completed three rounds of household surveys in 1994, 2005 and 2012.

The 1994 survey, also called the Human Development Profile of India (HDPI) collected data on 33,230 households. This survey included only the rural areas, although 277 households in 19 villages had been re-classified as towns by the 2001 census. HDPI can be accessed by following the steps laid out in this document: http://ihds.umd.edu/IHDS_HDPI_ConfidentialDataUseAgreement.pdf. STATA files and documentation for 33,230 households from this survey are available.

India Human Development Survey (IHDS), conducted in 2005, is a nationally representative survey of 41,554 households. Of the total, 26,734 households reside in rural areas and 14,820 in towns and cities. 2005 data can be downloaded at <http://www.icpsr.umich.edu/icpsrweb/>

DSDR/studies/22626. This survey contains a variety of information at different levels. An interview with a knowledgeable informant – typically the head of the household – covers socio-economic condition of the household including income, employment, educational status and consumption expenditure. An interview with an ever-married woman aged 15-49 covers health, education, fertility, family planning, marriage, and gender relations in the household and community. Short reading, writing, and arithmetic knowledge tests were administered to all available children aged 8-11 in the household. These tests were developed in collaboration with researchers from PRATHAM, India, and were pretested to ensure comparability across languages. Height and weight measurement was taken of children under age 5, aged 8-11, and their mothers. There was facilities assessment of one government and one private primary school and primary health care facility in the community. Village questionnaires assessed employment opportunities and infrastructure facilities in the village (Desai, et al. 2007).

The 1994 and 2005 surveys are partly linked, with 13,081 households common to the two surveys. A file containing identification variables for 1993-94 and 2004-2005 surveys, which permits the data from the two surveys for common households to be linked, will also be made available (<http://ihds.umd.edu/panel.html>).

The households interviewed in 2005 were re-interviewed in 2012. These data will be available in the public domain only in 2015 and will therefore not be available for our project.

The importance of the NCAER surveys in the context of the India-Brazil inequality project is that these are the only national level surveys that report income at the household level. These data will therefore be of critical importance for the analysis of income inequality. They also provide us with a link between income and consumption – so making it possible to generate an approximate estimate of income inequality on the basis of consumption data from other surveys. The NCAER surveys can also be used to study labour market behaviour with a rich set of control variables from a variety of domains such as fertility, education levels, health, and infrastructure at the village level. Obtaining such a wide variety of information from a single household is rare. This richness will increase the scope of analysis beyond labour markets, and will enable us to examine the impact on labour market outcomes of non market dimensions such as caste, fertility, or local infrastructure. Some of these varied and diverse associations cannot be analysed with the NSS surveys. Further, this is the only national source of household level panel data. Panel datasets are very useful in addressing self-selection and endogeneity issues in econometric analysis, and can help to produce unbiased estimates of determinants of income or consumption levels.

It is important to note that NCAER datasets are representative only at the national level, not at state or any lower regional level. Care should be taken in using the panel component of the surveys because the method of income calculation is different in the two rounds. Also, the panel component in 2004-05 survey only concerns rural households and there are concerns in using partial panel data. It is not clear how many out-migrating rural households and individuals have been picked up in the second survey. NCAER also notes that careful documentation is not available for the 1994 data. Neither can staff support be provided. Further there is some doubt about the accuracy of the survey. The questionnaires

are long and there are multiple questionnaires per household (see the link for the 2005 data set for this). There might be recall errors, survey fatigue, etc. which ends up reducing the quality of the data. Of course such problems may be present for any survey, but these survey questionnaires are particularly long.

3. Overall Comparability

The surveys of industry in Brazil and India are comparable in scope and content, making comparative analysis relatively straightforward. However, the range of industrial sectors covered in Brazil is distinctly wider. In addition, a much larger proportion of the Brazilian economy is covered because of the small share of the organized sector in India. A full analysis at the enterprise level would have to rely on the NSS surveys in India, of which there are too few to get a good time series.

It may well be worth making a specific comparison of the informal sector surveys in the two countries, even without analysis of changes over time. Closer review will be required to assess whether the IBGE and NSS surveys can be compared.

There is no counterpart to the NCAER survey in Brazil. However, in some respects the PNAD, which is wider than the NSS, is intermediate in scope between the NSS and the NCAER surveys. In any case the NCAER survey will enrich the analysis in India and make it possible to make direct comparisons of income inequality, even if it is only for one year (2004-05).

V. ADMINISTRATIVE DATA

5. Brazil

(i) RAIS

The Annual Report of Social Information (*Relação Anual de Informações Sociais* – RAIS) is an administrative register managed by the Ministry of Labour and Employment (MTE) and created by Decree 76.900 of December 23, 1975, in which all formal companies located in Brazil are required, on an annual basis, to declare information on staff they employed during the base year.

RAIS is therefore an annual picture of the formal labour market in Brazil, including permanent employees, statutory, temporary and fixed-term workers all of these categories with registered contracts. The 2011 RAIS contains information on 7.6 million establishments that declared 70.971 million employees in the year, of which 46.3 million were active in December – that is, they were still in employment at the end of the year.

RAIS database presents information on Brazil's formal workers such as: Sex, Race/colour; Schooling; Date of Birth; Nationality; Disability; Type of contract; Date of hiring and/or dismissal; Occupation according to CBO code (*Classificação Brasileira de Ocupações* – Brazilian Classification of Occupations); Contractual paid hours, Salary at hiring and dismissal/resignation; Earnings; Date of hiring and/or dismissal; Overtime; Time in employment; Type of dismissal.

As well as collecting information on workers RAIS also collects information on establishments such as: Economic activity (CNAE); and Size of the establishment, according

to the number of employees. All information has a nationwide reach and can be detailed down to the municipal level.

It is worth stressing that the information identifying the individuals made it possible for the MTE to create a derived database that allows tracking the work history of formal workers in the formal labour market. This is the RAIS Migra, where it is possible to visualize a worker's initial and final position each year, thus supplying tools for carrying out studies on occupational mobility, time in occupation, re-entering the formal labour market etc. The time that a worker has spent in informal activities cannot be collected by RAIS or RAIS Migra. The characteristics of RAIS Migra are generally similar to those of RAIS, only the manner in which the data are arranged and tabulated are different.

(ii) General Registry of Employed and Unemployed

The General Registry of Employed and Unemployed (*Cadastro Geral de Empregados e Desempregados – CAGED*) was established by Law 4.923 of December 23, 1965, which means that each employer is required to declare on a monthly basis all of the recruitments, dismissals and transfers of workers employed in his/her establishments according to the CLT system (*Consolidação das Leis do Trabalho – Consolidation of Brazilian Labour Laws*).

While RAIS is an annual picture of the formal labour market, CAGED allows one to keep up with the monthly flow of part of the formal workers in Brazil (those employed according to the CLT system). It is possible, for example, to verify which economic sectors and/or geographical regions had had a significant number of recruitments/dismissals and the absolute and relative variation of formal work, among other relevant information.

Regarding the information available, both RAIS and CAGED collect similar information, but CAGED doesn't have information on nationality, type of contract, total earnings, type of dismissal and overtime.

The access to RAIS, RAIS Migra and CAGED micro-data can be made by two alternatives: *i*) installing software X-OLAP and the databases in a local computer; and *ii*) online via the Program of Dissemination of Labour Statistics (*Programa de Disseminação das Estatísticas do Trabalho – PDET*), available at: <http://portal.mte.gov.br/portal-pdet/>. Both require prior MTE authorisation.

In the case of administrative registration the micro-data allow us to make any type of cross-tabulation with the information available, and some unique pieces of information from the administrative registry can be used as proxies in other databases, especially the PNAD.

The limitation is that they consider only workers in the formal labour market, that is, workers under registered contracts according to labour laws. Therefore, all workers without a registered contract, self-employed, entrepreneurs, domestic servants or unpaid family workers are not included. Furthermore, another inherent problem is the imputation errors due to the being registers declared by the companies themselves, who can add information without due care for its veracity.²⁰

2. India

An official register of all formal sector workers is unavailable for India. The closest to a register could be the Employment Market Information (EMI), a database based on returns filed by firms to employment exchanges. But unfortunately, the EMI does not have any information on individual employees (like for instance RAIS) but only a count of employees in the firm level, disaggregated by gender and occupational categories at the firm level. Reports based on the EMI data are readily available for download in the publications section of the website of Directorate General of Employment and Training (DGET) - <http://dget.gov.in/publications/welcome.html>. Micro data are not however readily available and need to be specifically requested from the Ministry of Labour. Data on worker flows (similar to CAGED) are also not available

3. Overall Comparability

Since there are no data for India that are directly comparable with RAIS and CAGED in Brazil, it does not appear possible to undertake any comparative analysis usual administrative records.

VI. MACRO LEVEL DATA

In addition to the mainly micro data discussed above, there are series of macroeconomic data in both countries which will be of value for the study of inequality. Some of the more important macroeconomic series are briefly described below.

1. Brazil

(i) National Accounts – Yearly and Quarterly

The System of National Accounts (SNA) is the main source of information concerning the flow of supply and demand of goods and services, as well as the generation and use of income in Brazil, and therefore of economic activity in the country. It is calculated on a quarterly basis by the IBGE according to the recommendations of the UN System of National Accounts 1993 (SNA93) manual, and the results are published in series of volumes and values at current prices and at the previous year's prices.

The historical series of Brazil's Gross Domestic Product (GDP) is available from 1900 and the current model of National Accounts refers to the year 2000, and is back-calculated to 1995.²¹

The information made available by the National Accounts for each period – quarter or year – are the Gross Domestic Product (GDP); *per capita* income; composition of the aggregated supply and demand; investment and savings rates; tax burden evolution; accumulation of capital; current transactions with the rest of the world; among others.

(ii) Regional Accounts and Gross Domestic Product of Municipalities.

With the aim of meeting regional demands and specificities, the IBGE carried out the first estimates of regional participation in the total GDP and *per capita* income by unit of the

Brazilian Federation, based on the economic censuses of 1970, 1975, 1980 and 1985. Only in 1996 did a debate begin on the question of a methodology specifically for the Regional Accounts that could be integrated with the National Accounts, something that was put into practice in the beginning of the 2000s.

The Regional Accounts are currently calculated from information on the gross production value, intermediate consumption and the gross added value of the Quarterly System of National Accounts. The historical series available for the Regional Accounts includes the period from 1995 to 2010, and can be detailed by sections such as units of the Brazilian Federation and economic activity for the following information: GDP; GDP *per capita*; and gross added value.

In its turn, the Municipalities' Gross Domestic Product aims to present information with the greatest amount of geographical detail possible and its calculation is made from the Regional Accounts. The Municipalities' GDP series currently includes the years from 1999 to 2010, making the following information available at current prices: GDP; *per capita* GDP; gross added value disaggregated by sector: agricultural/stock-rearing, industry, services, public administration; and taxes over products net of subsidies.

(iii) Relevance of the SNA for the Study

The fact that the National Accounts disaggregate the structure of the GDP both by production sector and by demand components allows the information to be used in complementary studies such as analyses of the functional distribution of income, that is, of the relative share of income from labour and from capital in the total national income. Also, Bastos (2012) shows that it is possible to use the Quarterly National Accounts in conjunction with the PNAD to create estimates for the years for which annual national accounts are still not available.

Although they include a smaller number of variables compared to the SNA, the Regional Accounts and GDP of Municipalities are important as they provide important economic information about different states and regions, which allows the calculation of regional indicators such as, for example, the HDI and the Gini index.

2. India

National accounts aggregates such as Net and Gross Domestic Products (GDP and NDP), consumption expenditure, capital formation; saving and capital stock are collected by the Central Statistical Organization of India. Such aggregates are reported for the economy as a whole as well as for the Public Sector, both at current and constant prices (with a particular base year) and published in the annual publication called National Accounts Statistics (NAS). The earliest of such national aggregate data are available as back series from 1951 to 2004-05, using 2004-05 as a base year. State Domestic Product series at the level of industry of origin are available from 1980-81 onwards. The most recent publication is National Accounts Statistics (NAS) 2012, which includes the Quick Estimates of macro-economic aggregates for the year 2010-11, advance Estimates of National Income for the year 2011-12 released on 7th February 2012 and Quarterly estimates of Gross Domestic Product (GDP) for the

years 2004-05 to 2010-11. All of these aggregate data are freely available on the website of Central Statistical Organization. These datasets will be useful to analyse trends inequality in terms of lagging states/regions within the country as well as lagging industries. Data and the related documentation can be accessed at Central Statistical Organization (2013).

3. Overall Comparability

In principle there is a high degree of comparability between Indian and Brazilian data at this level, because both countries follow internationally agreed standards. A comparison of regional accounts may be more problematic.

V. REFLECTIONS

Overall it is apparent that the data sources in the two countries are far from closely aligned. In some respects this is a historical accident, in that the statistical institutions that have developed in the two countries have acquired different characteristics, leading to different statistical strategies. Once a particular path is taken, it is difficult to change it, because both the Central Statistical Organization in India and the IBGE in Brazil are large, bureaucratic organizations with their own internal interests and momentums. Once the investment has been made in designing a particular statistical instrument, it becomes costly to change it.

But the differences in statistics between the two countries also reflect their realities. Labour market categories in Brazil reflect the dynamics of an urban labour market much more than the case in India, simply because the labour market in Brazil has been primarily urban for decades, while India is still mainly rural. On the other hand, the National Sample Survey in India has been used to obtain information on a wider range of topics than the PNAD in Brazil, reflecting national perceptions of statistical needs for policy and analysis. The focus on consumption in India and on income in Brazil also reflects different perspectives on the best ways to measure welfare. Since poverty has largely been seen as a question of low consumption in India, and income in rural areas is in any case hard to capture accurately with its large components in kind, the NSS has focussed exclusively on consumption. In Brazil, on the other hand, income has been considered sufficiently important that it is asked in the census (which is true of few other countries).

Some of the data categories are nevertheless reasonably comparable between the two countries, especially where there has been some international standardization through the UN system. This is true of national accounts statistics, and to a lesser extent demographic census data. The surveys of industry too are broadly comparable in design and concept (though wider in scope in Brazil). Some labour market measures are reasonably comparable (such as concepts of economic activity, occupation and industry), since these too draw on standard international concepts. But many of the more detailed breakdowns of the labour market and of the quality or security of employment are not comparable. The same is true of unemployment, for which the measures used in India are particularly idiosyncratic.

So much of the comparison between the two countries will be subject to qualification and approximation, and will require some statistical acrobatics. Comparing consumption

in India with income in Brazil will be particularly tricky. Inequality of land ownership in India, dominated by a peasant economy, cannot be directly compared with land ownership patterns in Brazil with its extensive commercial farming and much lower rural population density. It may be easier to compare exclusion from the labour market in the two countries than the quality of inclusion for those in work.

This does not make comparison impossible – it just means that interpretation will require a great deal of caution. The regular surveys in the two countries, the NSS and the PNAD, do provide reasonably comparable information on wages and employment, and this is a good starting point for much analysis of labour market inequality. Wage shares, gender differentials, regional differentials, industry differentials and much else can be compared even if the data are not identical. After all, if we want to examine inequality in labour market exclusion, as reflected in open unemployment or low labour force participation, this has to be done with the concepts which are relevant in the two economies concerned, not with some variable that is statistically comparable but not meaningful.

It is also important to bear in mind that much analysis has a large qualitative component anyway. When we study the long term institutional changes in each country, and their role in the changing patterns of inequality, the statistical data in each country is only one element of a wider, more qualitative analysis. It may be necessary to look at the impact on inequality of national politics, or of the role of trade unions, or of minimum wage policy – and in each case, the mechanisms involved may be subtle, and the variables needed to capture the outcomes different in the two countries. In this case, it is more important that the national data base be well adapted to understand what is happening at the national level, than achieving strict international comparability.

So the analytical strategy around this data base can be summarized in three components:

- For a limited range of variables that are very similar in the two countries, identical relationships can be explored in both
- For a large range of variables that cover similar ground, but are not identical in the two countries, parallel analyses can nevertheless provide a base for comparison
- And for a broader, more qualitative understanding, the outcomes and trends in both countries need a solid empirical foundation which reflects the national realities – and comparisons will be at the level of process and outcome, rather than at the level of statistical categories.

Notes

1. A proper census was carried out in 1872, when 20% of the population were still slaves. During the first Republican period (1889-1930), there were also several official censuses: in 1890, 1900 and 1920. However, it was only in 1936 that a specialized, dedicated body was created for geographic and social research – the IBGE – which has since been carrying out the census operations.
2. There are, however, two comments that must be made regarding the 1960 census. The main one concerns the absence of sample weights, which decreases the possibility of expanding the results of the sample to levels more disaggregated than a unit of the Brazilian Federation. The second factor refers to the fact that the “income” variable is only available in categorical form (income bands) and not as a continuous variable.

3. In the 1960, 1970 and 1991 Censuses, the condition of activity and of occupation referred to an individual's status in relation to his/her "habitual occupation" or main occupation carried out during the census year. In the other Censuses, there is clearer temporal delimitation: the question always refers to the week the questionnaire was applied. Therefore, in such cases, if the person did not work or seek employment during that time period, he/she is considered inactive.
4. Only in the last Census (2010) was it not possible to treat the schooling variable in terms of years of schooling, but solely in categories. IPUMS has also proposed forms of standardization and categorization of educational levels that are internationally comparable.
5. See J Krishnamurty, "Changing Concepts of Work in the Indian Censuses: 1901-61" *Indian Economic & Social History Review*, July 1977 Volume 14: 323-340, and J. Krishnamurty, "Occupational Structure" in *Cambridge Economic History of India* (Vol.II) ed. Dharma Kumar.
6. The IBGE also produces the Monthly Employment Survey (Pesquisa Mensal de Emprego - PME), which has very detailed information on work and income and also tracks the seasonal cycles of the labour market. However, its scope is restricted to six metropolitan regions.
7. It should also be noted that there are discrepancies between aggregate consumption estimates from the NSS and Central Statistical Organization (CSO)'s National Account Statistics. Experts state that consumption from the NSS is underestimated, in comparison to the Gross Domestic product calculated by the CSO. This is important in the interpretation of consumption inequality using the NSS data. But approximate adjustments can be made to handle these differences.
8. State samples are systematically collected for Tamil Nadu, Delhi, Maharashtra, Karnataka, Punjab, Gujarat, and Haryana. We need to think whether it is worth collecting them. The state samples, in combination with the central sample, permit within-state analysis with better reliability.
9. Agricultural establishments are considered those that deal in agriculture, cattle-breeding, forestry, poultry-farming, aquaculture, silkworm-breeding and others.
10. It must be mentioned that the 2006 Census used a different conceptual approach for certain topics, so that the information is not directly comparable with the previous censuses and require a compatibility exercise.
11. The only companies considered are those that are active in the Central Register of Enterprises (Cadastro Geral de Empress - CEMPRE) and with at least 50 per cent of revenue from industrial activity, that is, activities in sections B (extractive industries) and C (manufacturing industry) of the National Classification of Economic Activities (Classificação Nacional de Atividade Econômicas - CNAE 2.0). Companies from the informal sector of the economy are therefore not included. It must be pointed out that the adoption of the CNAE 2.0 by the IBGE sector surveys only took place in 2008, which allowed a greater level of disaggregation of the activities; however, the direct comparability with the information on fields of activity of previous years was lost. On the other hand, it is possible to make approximations for these different data.
12. The reason for this differentiation is that only the companies that are part of the first group are included in the PIA-Produto.
13. The PAIC sample is selected from companies active in the CEMPRE and whose source of revenue is mainly from section F (construction) of the CNAE 2.0.
14. Commercial activity is considered the purchase and resale of goods, without those goods undergoing any significant change.
15. The survey only includes companies active in the CEMPRE and with revenue from activities in section G (trade; repair of automotive vehicles and motorcycles) of the CNAE 2.0. The following segments are not included in the PAC: "maintenance and repair services for vehicles and motorcycles" and "repair of personal and domestic objects".
16. In the units of the Brazilian Federation of the Northern region, the only companies researched are those located in state capitals, with the exception of the state of Pará, where the metropolitan region of Belém is considered.

17. PAS only includes companies active in the CEMPRE, but the classification of service activities is wider and involves several sections of the CNAE 2.0.
18. Much like the PAC, in the Northern region, the PAS gathers information only for companies located in the capitals.
19. National Industrial classification codes 01405 to 37 (Central Statistical Organization, 2004).
20. The Unorganized Services survey covers hotels and restaurants (Section H of National Industrial Classification 2004); transport, storage and communication (section I); financial intermediation (J); real estate, renting and business activities (K); education (M); health and social work (N) and other community, social and personal service activities (O). This survey does not include Transport via railways, air transport, transport via pipeline, Operating of real estate of self-owned residential buildings, Monetary intermediation (group 651 of NIC 2004), and Activities of trade unions (sub-class 91200), religious organizations (sub-class 91910) & political organizations (sub-class 91920). (National Sample Survey Organization, 2008).
21. This is one of the reasons for which the variables workers' race/colour and occupation are not disclosed. With the aim of analysing the proximity of the RAIS information with the estimates of formally registered workers in PNAD, Negri et al. (2001) compared the two databases and observed the convergence of the distribution by sex and age, as well as the dispersion around the average income. In this way, in spite of its limitations, the authors concluded that RAIS is a reliable and robust source of data for carrying out studies on the formal labour market in Brazil.
22. The back calculation of National Accounts is the reweighting carried out with the adoption of a new year of reference, incorporating the structural changes of the economic activities and, where possible, the conceptual alterations that may have taken place, so as to make the series more homogenous.

References

Brazil

- Barbosa, Rogério J.; Marschner, Murillo; Ferrari, Diogo; Silva, Patrick; Prates, Ian A.; Barone, Leonardo S. "Ciências Sociais e Informação Quantitativa no Brasil: Entrevista com Elza Berquó e Nelson do Valle Silva". *Novos Estudos Cebrap*, n.95, 2013.
- Bastos, Estêvão Kopschitz Xavier. Distribuição funcional da renda no Brasil: estimativas anuais e construção de uma série trimestral. *Texto para Discussão 1702*. Brasília: Ipea, 2012.
- Brasil. Decreto n.º 76.900, de 23 de dezembro de 1975, Institui a Relação Anual de Informações Sociais – RAIS e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 24 dez. 1975, seção I, p. 17046.
- Brasil. Lei n.º 4.923, de 23 de dezembro de 1965, Institui o cadastro permanente das admissões e dispensas de empregados, estabelece medidas contra o desemprego e de assistência aos desempregados, e dá outras providências. *Diário Oficial [da] República Federativa do Brasil*, Brasília, DF, 29 dez. 1965, seção I, p. 13537.
- Brasil. Ministério do Trabalho e Emprego. *Relação Anual de Informações Sociais*. Disponível em: <http://www.rais.gov.br/>.
- Brasil. Ministério do Trabalho e Emprego. *Cadastro Geral de Empregados e Desempregados*. Disponível em: <http://portal.mte.gov.br/caged/>.
- Cavenaghi, Suzana. "Informe técnico - A experiência do Brasil nos módulos de domicílio e família nos censos demográficos". CEPAL - Centro Latinoamericano y Caribeño de Demografía (CELADE) - División de Población. Seminario-taller "Los censos de 2010 y las condiciones de vida" Santiago, 7 al 9 de junio de 2010.
- Corseuil, Carlos Henrique; Foguel, Miguel Nathan. "Uma sugestão de deflatores para rendas obtidas a partir de algumas pesquisas domiciliares do IBGE". Rio de Janeiro: Ipea, 2002. (Texto para Discussão, n. 897).

- Fishlow, Albert. "Brazilian size distribution of income". *American Economic Review*, v. 62, n. 2, 1972.
- Hoffmann, Rodolfo; Duarte, José Carlos. "A distribuição da renda no Brasil". *Revista de Administração de Empresas*, v. 12, n. 2, 1972, p. 46-66.
- IBGE. *Banco de Metadados*. Disponível em: <http://www.metadados.ibge.gov.br/>.
- IBGE. *Economia Informal Urbana – 2003*. Disponível em: <http://www.ibge.gov.br/home/estatistica/economia/ecinf/2003/ecinf2003.pdf>.
- IBGE. *Pesquisa Anual da Indústria da Construção – 2010*. Disponível em: ftp://ftp.ibge.gov.br/Industria_da_Construcao/Pesquisa_Anual_da_Industria_da_Construcao/2010/paic2010.pdf.
- IBGE. *Pesquisa Anual de Comércio – 2010*. Disponível em: ftp://ftp.ibge.gov.br/Comercio_e_Servicos/Pesquisa_Anual_de_Comercio/2010/pac2010.pdf.
- IBGE. *Pesquisa Anual de Serviços – 2010*. Disponível em: ftp://ftp.ibge.gov.br/Comercio_e_Servicos/Pesquisa_Anual_de_Servicos/pas2010/pas2010.pdf.
- IBGE. *Pesquisa Industrial Anual – Empresa 2010*. Disponível em: ftp://ftp.ibge.gov.br/Industrias_Extrativas_e_de_Transformacao/Pesquisa_Industrial_Anual/Empresa2010/piaempresa2010.pdf.
- IBGE. *Pesquisa Industrial Anual – Produto 2010*. Disponível em: ftp://ftp.ibge.gov.br/Industrias_Extrativas_e_de_Transformacao/Pesquisa_Industrial_Anual/Produto2010/piaproduto2010.pdf.
- IBGE. *Produto Interno Bruto dos Municípios – 2010*. Disponível em: ftp://ftp.ibge.gov.br/Pib_Municipios/2010/pibmunic2010.pdf.
- IBGE. *Sistema de Contas Nacionais Trimestrais*. Metodologia da Pesquisa. Disponível em: ftp://ftp.ibge.gov.br/Contas_Nacionais/Contas_Nacionais_Trimestrais/Metodologia_da_Pesquisa/srmtrimestrais.pdf.
- IBGE. *Sistema de Contas Regionais*. Disponível em: ftp://ftp.ibge.gov.br/Contas_Regionalis/2010/contasregionais2010.pdf.
- Langoni, Carlos. *Distribuição de renda e desenvolvimento econômico no Brasil*. Rio de Janeiro: Expressão e Cultura, 1973.
- Nascimento Silva, Pedro; Pessoa, Djalma Galvão Carneiro; Lila, Maurício Franca. F. "Análise estatística de dados da PNAD: incorporando a estrutura do plano amostral". *Ciência & Saúde Coletiva*, 7(4):659-670, 2002.
- Negri, João Alberto de; Castro Paulo Furtado de; Souza, Natalia Ribeiro de; Arbache, Jorge Saba. Mercado formal de trabalho: comparação entre os microdados da RAIS e da PNAD. *Texto para Discussão 840*. Brasília: Ipea, 2001.
- Osorio, Rafael Guerreiro. "O sistema classificatório de 'cor ou raça' do IBGE". IPEA - *Texto para Discussão* N° 996, Brasília, novembro, 2003.
- Reis, Eustáquio José et al. "Áreas mínimas comparáveis para os períodos intercensitários de 1872 a 2000". *1º Simpósio Brasileiro de Cartografia Histórica*. Paraty, 10 a 14 de maio de 2011.
- Rigotti, José Irineu Rangel. Variáveis de educação dos censos demográficos brasileiros de 1960 a 2000. In: Rios-Neto, Eduardo Luís Gonçalves, Riani, Juliana Ruas (org.). *Introdução à Demografia da Educação*. Campinas, ABEP, 2004.
- Singer, Paul. *Dominação e desigualdade (Estrutura de Classes e Repartição da Renda no Brasil)*. Rio de Janeiro, Paz & Terra, 1981.
- India**
- Azam, Mehtul and Nishith Prakash. 2010. "A distributional analysis of the public-private wage differential in India". IZA DP No. 5132. August 2010.
- Central Statistical Organization, 2004. National Industrial Classification. Ministry of statistics and programme implementation, Government of India.
- Central Statistical Organization. 2013. Reports and publications, National accounts division. Can be found at http://mospi.nic.in/Mospi_New/admin/Login.aspx, last accessed on July 30, 2013.

- Das, Panchanan. 2012. "Wage Inequality in India: Decomposition by Sector, Gender and Activity Status". *Economic and Political Weekly*. December 15.
- Desai, Sonalde, Amaresh Dubey, B.L. Joshi, Mitali Sen, Shariff, Abusaleh, and Reeve Vanneman. India Human Development Survey (IHDS) [Computer file]. ICPSR22626-v1. University of Maryland and National Council of Applied Economic Research, New Delhi [producers], 2007. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-07-30. available online at http://ihds.umd.edu/IHDS_files/ihdsdoc01.pdf, last accessed on July 30, 2013.
- National Sample Survey Organization. 2004. Evolution of the Sample Design in NSS. Available at http://mospi.nic.in/Mospi_New/upload/nso/design_golden.pdf, last seen on July 30, 2013. Ministry of Statistics and program implementation, Government of India.
- National Sample Survey organisation. 2008. Service Sector in India (2006-07) Operational Characteristics of Enterprises. Ministry of Statistics and program implementation, Government of India.
- Office of the Registrar General and Census Commissioner. 2011. History of Census in India. Drop-in-Article on Census - No.5. Ministry of Home Affairs, Government of India.
- Sarkar, S and Balwant Singh Mehta. 2007. "Growth of Employment and Earnings in Tertiary Sector, 1983-2000". *Economic and Political Weekly*, March 17.

**IHD WORKING PAPER SERIES
(NEW SERIES)**

<i>No.</i>	<i>Authors</i>	<i>Title</i>
WP 03/2014	Alexandre de Freitas Barbosa, Maria Cristina Cacciamali, Gerry Rodgers, Vidhya Soundarajan, Fabio Tatei, Rogerio Barbosa, J. Krishnamurty	Data Sources for the Analysis of Labour Market Inequality in Brazil and India (CEBRAP-IHD) Research Project on Labour Market Inequality in Brazil and India
WP 02/2014	Sheila Bhalla	Scarce Land: Issues, Evidence and Impact
WP 01/2014	Ajit K. Ghose	India's Services-Led Growth
WP 06/2013	Gerry Rodgers, Sunil K. Mishra and Alakh N. Sharma	Four Decades of Village Studies and Surveys in Bihar
WP 05/2013	Sumit Mazumdar Alakh N. Sharma	Poverty and Social Protection in Urban India : Targeting Efficiency and Poverty Impacts of the Targeted Public Distribution System
WP 04/2013	Dev Nathan, Govind Kelkar and Shivani Satija	Witches: Through Changing Contexts Women Remain the Target
WP 03/2013	Shivani Satija	Violence Against Women in Rural Bihar- A case of four villages
WP 02/2013	Sheila Bhalla	Behind the Post 'Challenge to The Functional Efficiency of India's Established Statistical Institutions
WP 01/2013	Preet Rustagi, Dev Nathan, Amrita Datta and Ann George	Women and Work in South Asia: Changes and Challenges
WP 05/2012	Amrita Datta, Gerry Rodgers, Janine Rodgers and B.K.N. Singh	A Tale of Two Villages: Contrasts in Development in Bihar
WP 04/12	Janine Rodgers	Labour Force Participation in Rural Bihar: A Thirty-year Perspective based on Village Surveys
WP 03/12	K.P. Kannan	How Inclusive is Inclusive Growth in India?
WP 02/12	Sheila Bhalla	Notes on Land, Long Run Food Security and the Agrarian Crisis in India

WP 01/2012	Gerry Rodgers	Understanding Unequal Economic and Social Outcomes in Rural Bihar: The Importance of Caste, Class and Landholding
WP 03/2011	Dev Nathan and Sandip Sarkar	Global Inequality, Rising Powers and Labour Standards
WP 02/2011	Preet Rustagi and Rajini Menon	Gender Asset Gaps and Land Rights in the Context of the Asia-Pacific Region
WP 01/2011	D. Narasimha Reddy	NREGS and Indian Agriculture: Opportunities and Challenges
WP 03/2010	Gerry Rodgers and Janine Rodgers	Inclusion or Exclusion on the Periphery? Rural Bihar in India's Economic Growth
WP 02/2010	R Radhakrishna, C Ravi and B Sambhi Reddy	Can We Really Measure Poverty and Identify Poor When Poverty Encompasses Multiple Deprivations?
WP 01/2010	Alakh N. Sharma	Political Economy of Poverty in Bihar: Nature, Dimensions and Linkages
WP 03/2009	Dev Nathan and Govind Kelkar	Markets, Technology and Agency: Indigenous People and Change
WP 02/2009	Aseem Prakash	Towards Understanding the Nature of Indian State and the Role of Middle Class
WP 01/2009	C Upendranadh and Rukmini Tankha	Institutional and Governance Challenges in the Social Protection: Designing Implementation Models for the Right to Work Programme in India
WP 06/2008	Dipak Mazumdar and Sandip Sarkar	The Employment Problem in India and the Phenomenon of the Missing Middle
WP 05/2008	Aseem Prakash	Social Conflict, Development and NGOs: An Ethnographic Study
WP 04/2008	Balwant Singh Mehta and Kerren Sherry	Wages and Productivity of Child Labour: A Case Study of Zardosi Industry
WP 03/2008	Sheila Bhalla	Scarce Land: The Cases of All India and West Bengal
WP 02/2008	T.S. Papola and R.P. Mamgain	Market Access to Small Primary Producers: A Study of Vegetable Growers in the Supply Chain
WP 01/2008	Preet Rustagi	Rural Child Labour Markets in India: Nature of Child Work Participation and Role of the Family

WP 03/2007	Alakh N Sharma	Flexibility, Employment and Labour Market Reforms in India
WP 02/2007	Preet Rustagi	Rural Child Work, Labour and Daily Practices: A Time Use Survey-based Analysis
WP 01/2007	R.P. Mamgain	Growth, Poverty and Employment in Uttarakhand
WP 04/2006	Sheila Bhalla	Common Issues and Common Concerns in the SAARC Region: Employment Generation and Poverty Reduction
WP 03/2006	Rajendra P. Mamgain and Balwant Singh Mehta	Employment and Earnings in Uttaranchal: Trends and Policy Issues
WP 02/2006	Preet Rustagi	Women and Poverty: Rural-Urban Dimensions
WP 01/2006	Dipak Mazumdar and Sandip Sarkar	Growth of Employment and Earnings in the Tertiary Sector, 1983-2000
WP 05/2005	Sheila Bhalla	Recent Developments in the Unorganised Rural Non-Farm Sector
WP 04/2005	Preet Rustagi	Challenges for Economic Empowerment of Women In South Asia
WP 03/2005	Alakh N Sharma	Agrarian Relations and Socio-economic Change in Bihar
WP 02/2005	Preet Rustagi	The Deprived, Discriminated and Damned Girl Child: Story of Declining Child Sex Ratios in India
WP 01/2005	Dipak Mazumdar and Sandip Sarkar	Agricultural Productivity, Off-Farm Employment and Rural Poverty: An Analysis Based on NSS Regions

The Institute for Human Development (IHD), a leading centre for studies on labour markets, livelihoods and human development, aims to contribute to the building of a society that fosters and values an inclusive social, economic and political system, free from poverty and deprivations. Towards achieving its goal, it engages in analytical and policy research, teaching and training, academic and policy debates, networking with other institutions and stakeholders, and publication and dissemination of the results of its activities. The major themes of current work of IHD are: poverty, inequality and well-being; labour markets and employment; social protection; women and children; marginalized social and economic groups and lagging regions; and governance and institutions for human development.



Institute for Human Development

NIDM Building, IIPA Campus, IP Estate, New Delhi - 110002

Phones: + 91-11-23358166/23321610 • Fax: +91-11-23765410

Email: mail@ihdindia.org • Website: www.ihdindia.org