Data Appendix for "Beyond GDP?
Welfare across Countries and Time"

Charles I. Jones

Stanford GSB and NBER

Peter J. Klenow

Stanford University and NBER

September 1, 2010 – Version 2.0

1. Introduction

This data appendix has two parts, one for the macro data used in the main part of the paper and one for the micro data used at the end of the paper.

2. Macro Data

The basic data sources that we use and an overview of the manipulations of this data are described in the main paper in Section 3.1. Some of the basic underlying data that goes into our calculations is available in the spreadsheet where we report our extended results; a link to that file is available here.

Once we've double-checked and better annotated our matlab and stata programs, these will be made available. If you'd really like to see the programs immediately, send us an email and we'll be happy to provide them.

2.1. Main Matlab Programs

MakeData9.m: This is the main program for loading the various data sources that are used in the main calculations. The parallel program for the growth rate analysis is MakeDataGrowth9.m.

Rawls10.m, Growth9.m: These are the main programs for computing the results. Rawls10 computes the basic results displayed in Tables 1 and 2 and Figure 3. Growth9 does the same thing for the growth results in Tables 3 and 4 and Figure 4.

PlotBasicData9.m: This program generates basic plots of the underlying data, including Figures 1 and 2.

Robustness checks: The following programs generate the results in the robustness check tables (in addition to Rawls10.m and Growth9.m):

- Rawls10ev, Growth9ev

- Rawls10cv, Growth9cv

- Crra10c0, Growth9c0

- Crra10cbar, Growth9cbar

- Rawls10France, Growth9France

- Rawls10KY, MakeData9KY.m.

Subsidiary programs: These main programs call several functions:

- Getubar10.m: calculate the value of $\bar{u}$ to match the U.S. value of life for a 40-year old in 2000, as described in the paper.

- regions63.m: region variables, based on PWT6.3.

- A bunch of smaller matlab functions , such as weightedaverage.m, cshow.m, etc.

2.2.   Programs for Reading and Cleaning Data

WIID2cXCRatio63.m, WIID2cGrowth63.m: These programs read and clean the WIID data on Gini coefficients. The source for our inequality data is the UNU-WIDER World Income Inequality Database, Version 2.0c, dated May 2008. The WIID database reports income and consumption Gini coefficients from a variety of micro data sets for many countries and years. We use consumption measures when they are available and infer consumption measures from income measures when only the latter are available. For the cross-sectional analysis, we average across available observations that meet a certain quality threshhold for the period 1990 to 2006. For the

time-series analysis, we use data from 1974–1986 to construct a 1980 estimate and from 1994–2006 to construct a 2000 estimate.

LifeExpectancyWB63.m: Loads the life expectancy data from the World Bank. These data are taken directly from the World Bank's HNPStats database.[1]

WBAdultPopulation.m: We measure time spent in leisure or home production as the difference between a time endowment and time spent in employment. Our measure of time engaged in market work aims to capture both the extensive and intensive margins. For the extensive margin, the Penn World Tables, Version 6.3 provides a measure of employment, apparently taken from the Groningen Growth and Development Center. We divide this employment measure by the adult population, i.e. those ages 15 and over (obtained from the World Bank). The program WBAdultPopulation.m loads the fraction of the population aged 15 and over.

AnnualHours2.m, AnnualHoursGrowth.m, ILOWeeklyHoursGrowth.m, ILOWeeklyHours.m, OECDhours.m: Our measure of the intensive margin is annual hours worked per worker. For OECD countries, this measure comes directly from SourceOECD. For non-OECD countries, we impute annual hours per worker using a measure of average weekly hours in manufacturing from the International Labour Office, based on a simple regression for the 28 countries that have data on both variables. For countries without weekly hours or OECD data, we assign a sample mean to the intensive margin. In particular, for countries richer than 50% of the U.S. (basically Portugal's income level) that are missing data, we use the average annual hours for countries above this income level; ditto for countries that are poorer than 50% of the U.S. value. These two averages are 1756 hours, a value that is assigned to 8 "rich" countries and 1944 hours, the value assigned to 94 "poor" countries.

---

[1]See http://go.worldbank.org/N2N84RDV00, series code SP.DYN.LE00.IN.

3.   Micro Data

3.1.   Overview

For the Household Survey data, we wrote two Stata programs to analyze the data for each country-year:

- WBC_YR_sumstats.do

- WBC_YR_lamstats.do

WBC refers to the three-letter World Bank Country Code (FRA, IND, MEX, ZAF, or USA). YR refers to the year of the survey (e.g. 06 for 2006, 85 for 1985).

The "sumstats"files create datasets WBC_YR.dta with the following common set of variables for each individual covered in that household survey:

- hhid (household id code)

- hhsize (number of individuals in the household)

- age (age of the individual)

- hhexp (total household expenditures on nondurables and services)

- leisure (fraction of the time endowment the individual is not working)

- weight (sampling weight)

See the text for the number of individual observations in each country-year. Below we describe in more detail how we constructed household expenditures and hours worked for individuals in each survey. In each case, we define expenditures as those on nondurables and services to the exclusion of durable goods. We divide household expenditures by the number of individuals in the household to obtain individual consumption. And we define leisure as the proportion of total hours in a year that a person does not work:

$$leisure = \frac{(5840 - \text{annual hours worked})}{5840},$$

where 5840 = 365 days · 16 waking hours per day.

The "lamstats" files read in WBC_YR.dta and calculate welfare relative to the U.S. in the same year (in log terms) and its additive components: terms due to life expectancy, average consumption, consumption inequality, average leisure, and leisure inequality. These calculations are made using sampling weights.

3.2. France

For France we use the Family Budget Survey (Enquete Budget de Famille - or EBF) conducted by the National Institute for Statistics and Economic Studies. The EBF contains information about each surveyed household's income, expenditures on final consumer goods, and demographic characteristics. The survey is a repeated cross section, conducted every five years since 1979, and is representative at the national level once sampling weights are applied. We use the earliest and latest years with the necessary data, 1984 and 2005, to calculate growth rates. We use the latest year, 2005, for comparison with the U.S.

We construct consumption by adding up the reported expenditures on the following: food, clothing, housing (e.g. rent and estimated rent for those who own their house, payments for electricity, etc), accessories for the house (e.g., dishes, cooking utensils, light bulbs) excluding furniture, medical services, communication services (e.g., postal and telephone services), leisure and cultural spending (expenditures on pets, gardening, theater tickets, and entertainment events), accessories for personal care (soaps, perfumes, brush, etc.), professional services (lawyers, accountants, funerals, etc.), and transportation services (bus or subway tickets, taxi fees, etc.). In these categories we include estimated home production for self-consumption and gifts received. In each case we exclude durables (furniture, durable leisure goods, vehicles, etc.). Besides excluding the value of expenditures on durables, we exclude the following: maintenance; repair and expansion of housing; deposits in savings accounts; loans and debt payments; income and property taxes; purchases of land, houses or condominiums; transfers made to acquaintances and for charity; and value of items stolen.

The dataset contains information about weekly hours worked for all members

of the household aged 16 years and older. Because the household survey does not ask about weeks worked in the year, we use the OECD statistics for weeks worked per worker in France: 43.5 weeks in 1984 and 41.0 weeks in 2005. For those in the household under 16 years old, we assume zero hours worked so that their fraction of leisure time is 1.

3.3.   India

For India we use the National Sample Survey (NSS) conducted by the Indian Ministry of Statistics and Programme Implementation. In some rounds (years) the NSS contains information about expenditures on final consumer goods and demographic characteristics for a cross-section of households. We use the earliest and latest years with the necessary data, 1983-1984 and 2004-2005, to calculate growth rates. We use the latest year, 2004-2005, for comparison with the U.S.

The 1983-1984 survey has separate schedules on consumption and time use, respectively. We were able to plausibly match roughly one-half of the households in the two schedules. The relevant statistics (consumption and leisure levels and inequality by age) are similar whether we use only the matched households or all households. As the utility function we use in the micro calculations is additively separable in consumption and leisure, we simply calculated the consumption and leisure terms on the separate samples in 1983-1984. The 2004-2005 survey covered both consumption and leisure for a common set of households.

Our measure of consumption includes expenditures on the following, which the survey asks about specifically: food (itemized), fuel and light, cinema/theater/video, tuition fees, newspapers/magazines/fiction, medical expenses, toilet articles, regular (commuting type) and other journeys, house rent, clothing and footwear. Respondents were asked to include the value of home production consumed in these categories. We exclude spending on durable goods, which were also asked about and itemized (furniture, appliances, etc.).

The dataset contains "daily time disposition" for the prior seven days for each household member. For each day, two main activities were identified and recorded.

We count the following as working: self-employment, unpaid family labor, regular salary/wage employment, and casual wage labor. The survey asks the respondent to assign each activity a "full-intensity" (4 hours or more) or a "half-intensity" (1-4 hours). For constructing hours worked in a week, we treat full-intensity as 8 hours and half-intensity as 2.5 hours for the first 5 days of the week, and half these levels for days 6 and 7. With these values, many individuals work the resulting maximum of 48 hours in a week in the Indian survey. We multiply weekly hours worked by 52 weeks to get annual hours worked. As these conversions are admittedly arbitrary, however, the Indian leisure results must be taken with particular caution.

3.4. Mexico

For Mexico we use the National Survey of Household Income and Expenditure (Encuesta Nacional de Ingresos y Gastos de los Hogares - or ENIGH) conducted by the National Institute of Statistics and Geography. ENIGH contains information about each surveyed household's income, expenditures on final consumer goods, and demographic characteristics. The survey is a repeated cross section, conducted every two years, and it is representative at the national level once sampling weights are applied. We use the earliest and latest years with the necessary data, 1984 and 2002, to calculate growth rates. We use the latest year, 2002, for comparison with the U.S.

Our measure of consumption includes expenditures on the following: housing (rent and estimated rent for those who own their house), food, clothing and accessories, household services (e.g. utilities), accessories for the house (e.g., light bulbs) excluding furniture, leisure spending (expenditures on pets, gardening, and entertainment events), accessories for personal care (soaps, perfumes, brush, etc.), professional services (lawyers, accountants, funerals, etc.), and transportation services (bus or subway tickets, taxi fees, etc.). In these categories we include estimated home production for self-consumption and gifts received. In each case we exclude durables (furniture, durable leisure goods, vehicles, etc.). Besides excluding the value of expenditures on durables, we exclude the following: maintenance, repair and expansion of housing; deposits in savings accounts; loans and debt pay-

ments; income and property taxes; purchases of land, houses or condominiums; transfers made to acquaintances and for charity; and value of items stolen.

The dataset contains information about weekly hours worked (in the previous month) for all members of the household aged 12 years and older. Because the survey does not ask about weeks worked in the year, for both years we assume those working work 43 weeks, as in the OECD statistics for Mexico in 2002 (weeks worked are not available in the OECD statistics for Mexico for 1984). For those in the household under 12 years old, we assume zero hours worked so that the fraction of leisure time is 1 for them.

3.5.   South Africa

For South Africa we use the Integrated Household Survey (HIS) conducted by the South Africa Labour Development Research Unit (at the University of Cape Town) in collaboration with the World Bank. The HIS contains information about each surveyed household's income, expenditures on final consumer goods, and demographic characteristics. The survey is a single cross section, conducted from mid-1993 through early 1994.

Our measure of consumption includes expenditures on the following: housing (rent and estimated rent for those who own their house), utilities, food, personal items, clothing, health care, schooling, and transportation. The survey explicitly asks about food consumption from own production. We exclude durables expenditures (including home repairs).

The dataset contains information about monthly hours worked for all members of the household aged 16 years and older. Workers include the self-employed, those employed in a family business (including crop production for own consumption), those with regular employment, and those with casual or temporary employment. We multiply monthly hours worked by 12 months to get annual hours worked. For those in the household under 16 years old, we assume zero hours worked so that the fraction of leisure time is 1 for them.

3.6.   United States

For the U.S. we used the Consumer Expenditure Survey (CES) carried out by the U.S. Bureau of Labor Statistics. The CES contains information about each surveyed household's income, expenditures on final consumer goods, and demographic characteristics. The survey is a rotating panel of households, with each household reporting expenditures and hours worked for up to four consecutive quarters. We use various years from 1984 through 2006 for comparison with other countries (e.g. South Africa in 1993), and for calculating U.S. growth rates from 1984-2006.

We use a cleaned version of the CES data generously made available by Dirk Krueger and Fabrizio Perri, following their methodology in Krueger and Perri (2006). As a result we use their definitions of consumption expenditures and hours worked. Specifically, we start from their definition of nondurables (food, personal care, fuel, utilities, household operations, public transportation, apparel, education, reading, health services, and miscellaneous personal services) and add the following: services from vehicles, other vehicle expenses, services from owned primary residence, rent, other lodging expenditures, and entertainment. We exclude durables expenditures.

The CES contains information about weeks worked and hours per week for the respondent and for the respondent's spouse (if any). For other members of the household, we assumed zero hours worked so that the fraction of leisure time is 1.

References

Krueger, Dirk and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies,* January 2006, *73* (1), 163–193.