World Bank Center for Development Data (C4D2)

Partnership for Capacity Development in Household Surveys for Welfare Analysis

Measuring Consumption

Instructor Manual

This draft: May 21, 2019

PRELIMINARY

DO NOT CIRCULATE OR CITE WITHOUT THE AUTHORS' PERMISSION

This draft was prepared by Giulia Mancini and Giovanni Vecchi (University of Rome Tor Vergata). We would like to acknowledge the contribution provided by Nicola Amendola and Sédi-Anne Boukaka (University of Rome Tor Vergata) in preparing the material underlying the course. We are grateful to Gero Carletto, Michelle Jouvenal, Shelton Kanyanda and Alberto Zezza (World Bank) for helpful suggestions.

Introduction: how to use this Manual

This Manual is addressed to the Instructor of the course "Measuring Consumption", prepared by the World Bank's Center for Development Data (C4D2). It is part of a **Teaching Package** that also contains the following materials:

1. Course syllabus
2. Lecture slides (classroom version, PowerPoint, and handout version, pdf)
3. Reading package
4. Final exam and grading sheet
5. Course evaluation form

The Manual explains how best to make use of each of these items, and provides practical guidance to support the Instructor in the preparation and delivery of the course.

The rest of the Manual is organized as follows. First, the section titled **The Course** gives a general outline of the course, lays out its overall learning objectives, and explains how they will be reached, by providing a detailed Syllabus. This section also contains some important logistical information (required facilities, teaching materials, etc.), that should be carefully assessed in preparation for the course.

Next, the section titled **Teaching and Communication Principles** covers both some general principles of public speaking and communication, as well as some tips that are useful for the delivery of this particular course.

The remaining sections provide information for each of the **15 lectures** included in the course. For each lecture, you will find *(i)* learning objectives; *(ii)* a description of any preparatory work that the Instructor is advised to do before giving the lecture; *(iii)* a time allocation plan, which advises Instructors on how best to allocate the available classroom time; *(iv)* a writeup of the topics covered in the lecture, designed to support the Instructor's understanding of the slides; and *(v)* keys to any exercises associated with the lecture.

The Course

Target audience and pre-requisites

The target audience is Master students enrolled in Graduate Programs in Statistics, Official Statistics, and related fields, offered by the Regional Training Centers.

This course is designed to be self-contained, therefore it has no compulsory pre-requisites. However, it is an advantage if students have at least a basic command of Statistics, Sampling Design, and Microeconomics.

Learning objectives

The ultimate goal of the course is to improve the **quality** and **comparability** (both over time and across countries) of household survey data used for living standards measurement, by training students who will pursue careers within Statistical Offices and Institutions.

The course focuses on two main topics: first, it lays out the **conceptual framework** underlying the measurement of living standards; second, it offers practical **guidelines for survey design and data collection**, in the specific context of household consumption and expenditure modules.

Course syllabus

The course is organized in **15 lectures**. Teaching materials are optimized for each lecture to last two hours, including a short break of about 15 minutes.

A number of **readings** are associated to each lecture: all of them are provided in the form of a Reading Package, to be shared with students. Students should be encouraged to complete readings before each lecture. Then, classroom presentations provide guidance for the sometimes technical and challenging published material. A few selected readings are to be considered essential to a student's learning, and are marked **required**. All other readings are optional, but useful for those interested in knowing more.

Lectures also contain a number of **exercises**, to be assigned as homework. This Manual includes keys and evaluation guidelines for the exercises.

Plan of the lectures:

1) **Measuring living standards: a conceptual framework**
   From theory to practice: Income vs. Expenditure/Consumption vs. Wealth

2) **The Consumption Aggregate**
   Four building blocks: (i) food, (ii) non-durable non-food items, (iii) housing, and (iv) durable goods. Adjustments: for household size and needs, for within-year inflation.

3) **Understanding household surveys**
   Types of surveys; defining features of household surveys that measure consumption, as opposed to other statistical instruments.

4) **Principles of questionnaire design**
   Overview of the survey process; general principles of questionnaire design, as they apply to the measurement of consumption expenditures.

5) **Measuring food consumption: the foundations**
   Concepts (*e.g.* acquisition vs. consumption). Questionnaire design (recall vs. diary)

6) **Measuring food consumption: questionnaire design**
   Questionnaire design (list of food items, meal participation, seasonality)

7) **Food away from home and the use of non-standard units for measuring food consumption**
   Questionnaire design (non-standard measurement units, food away from home)

8) **Measuring consumption of non-durable non-food goods**
   Analytical needs; questionnaire design.

9) **Durable goods**
   Analytical needs and questionnaire design.

10) **Housing**
    Analytical needs; questionnaire design.

11) **Data validation and diagnostics**
    Key principles of statistical data editing.

12) **Outlier detection and treatment**
    Key principles of outlier detection and treatment.

13) **Measuring inequality**
    Lorenz curve and Gini index; other selected inequality measures.

14) **Measuring poverty**
    Poverty lines and measures.

15) **Describing data**
    Preparation of a “tabulation report”.

Course evaluation

The Teaching Package contains a **course evaluation form** designed to collect feedback from students. The evaluation form is anonymous, and should be handed out after the last lecture. Instructors are encouraged to share results with the C4D2 team, to help improve future iterations of the course.

Final exam

The final exam is a take-home assignment. The text of the final exam and a grading sheet are provided as part of the Teaching Package.

Logistics

The total amount of **classroom time** required for the course is 30 hours (there are 15 lectures, each lasting two hours). Training Centers are advised to schedule one or, at most, two of these lectures in one day. Because the final exam is a take-home assignment, no further classroom time is required after the end of the course.

The course calls for the use of PowerPoint presentations to support the Instructor's delivery: therefore, a room outfitted with a **projector** is required.

Instructors can use the following **checklist** to organize the logistics of the course:

Before registration:

- ❑ Share the syllabus of the course
- ❑ Confirm availability of required facilities (projector)

Before Lecture 1:

- ❑ Edit Lecture 1's "Contacts" slide, adding the Instructor's email address, office hours, etc.
- ❑ Print out slides (handout version) for each student
- ❑ Prepare to share digital copy of reading package with students

Before Lecture 15:

- ❑ Print out evaluation forms for each student
- ❑ Print out (or prepare to share digital copy of) final exam

Please note: printing out the **handout version** of the course slides, so that each student can use the printed slides for notes and exercises, is strongly encouraged.

Lecture 1
Measuring living standards: a conceptual framework

Learning objectives

The goal of this lecture is to provide the theoretical basis for the choice of a measure of living standards, which analysts need in order to estimate both poverty and inequality. The material covered in the lecture justifies the focus on a consumption-based measure of living standards.

Suggested preparation

The Instructor is assumed to be familiar with the material contained in section 2.2 ("Money-metric utility") of Deaton and Zaidi (2002), as well as with references therein listed.

Time allocation

Students' introductions (name, background)	**15 min**
Introduction	15 min
What is the standard of living?	30 min
Break	**15 min**
Choosing a measure of living standards	45 min

Annotated lecture

As the first lecture begins, it is good practice to give students a bird's-eye view of the whole course. This task does not require much time, but clarifying what the course is about, and why it is important – and hopefully interesting, too – is key. The aim is to convey **motivation** (students should feel they are about to learn important concepts and tools), and to clarify **expectations** from both the Instructor's and the students' sides. This is why the first bunch of slides gives an overview of the topics that will be covered in the following classes, and, even more importantly, explains why students should take the time to study them.

The broad, long-term objective of the course, that justifies an interest in the topics covered, is that of improving the quality of household survey data used for measuring living standards (which in turn are the basis for poverty and inequality measurement).

What do we mean by **data quality**, exactly? Before going into details, we can take this opportunity to make a general point: learning to use words and terms as precisely and technically as possible is part of the aims of this course. The attitude that we encourage in students is to ask questions such as "what do we mean by 'something', exactly?" where emphasis is on '**exactly'**. Answers are in the **literature**, which students will be encouraged to consult throughout the course, instead of relying entirely on 'common sense'. Here, we can consider engaging with students, and asking *them* "how would you define and assess 'data quality', exactly?". We allocate only a a couple of minutes to discussing answers, then move on.

Back to our question 'What do we mean by **data quality**, exactly?'. In their manual on survey quality, Biemer and Lyberg (2003: 13-18) review different ways in which data quality has been defined by various statistical institutions: some common criteria emerge. According to the first criterion, in order to be of good quality, data should be relevant for the needs of the users. Therefore, it is crucial for data producers to understand and clearly define the research objective that justifies their data collection efforts (what will the data be used for?). The remaining criteria can be met by following "best practices", which are based on scientific evidence and on the experience consolidated by international institutions.

The course is designed to help in meeting most of these quality criteria: its specific objectives are to provide a conceptual framework for the measurement of living standards (this addresses the relevance criterion); and to provide guidelines for survey design and data collection, tailored to the production of household consumption and expenditure data (this addresses all other quality criteria, particularly those of accuracy and comparability).

Because the course is focused on the collection of data for the purpose of inequality and poverty measurement, this slide explains the 5 building blocks that are needed to construct estimates of inequality and poverty:

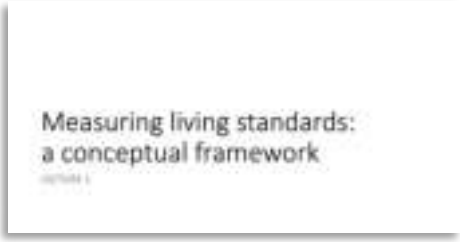
1) a **measure of living standards**, that is, a quantifiable concept of what living standards are and how they can be defined in practice;
2) high-quality **data** on households' living standards, typically collected through surveys;
3) a distribution of living standards, that is, the actual data that come from the field, and that describe the variation of living standards across the population of interest. At this stage, it is possible to produce **inequality** estimates;
4) a critical level, a **poverty line**, below which individuals are classified as "poor", that is, a value, a fixed amount of whatever quantity indicates individual living standards, that gives us a threshold below which individuals are classified as poor;
5) one or more **poverty measures**, such as a simple headcount of how many poor individuals there are, or any other more complex measure of our choice, to summarize the results of the analysis.

The focus of this course will be on the first 2 building blocks: a measure of living standards (again, this is the conceptual foundation of the analysis, which must be embedded in the

design of the survey); and high-quality data that describe households' living standards. This is a good time to comment on the importance of '**theory**': the course will combine 'theory' (first building block) with 'practice' (second building block). There is no such a thing as good data collection in the absence of a theoretical framework. Building blocks 3) to 5) venture into analytical territory, and are complex enough to be the subject of a different course.

This slide gives an overview of the topics to be covered in each of the 15 lectures of the course. We should not read and comment item by item – this would take too long, and probably bore students to death. We simply remind students that there is a Syllabus, and encourage them to check it and use it as a roadmap throughout the duration of the course.

Next, some practical instructions that are of interest for students. The **Instructor's contacts** should be personalized.



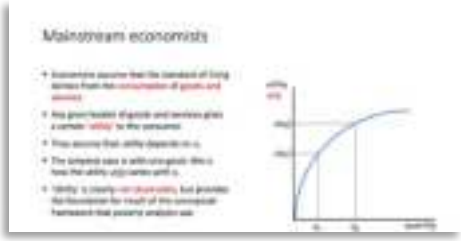
After the introduction, we enter into the actual contents of lecture 1.

The first part of the lecture answers the question: “What is the standard of living?”. Here, the Instructor can involve students, and encourage them to give their own answers to the question. The point that should emerge from this short discussion is that the answer is highly complex, highly subjective, and in no way easy to agree on. We emphasize that is good to keep this **conceptual complexity** in mind, although statisticians, economists, data analysts are interested in providing a quantifiable answer – one that allows for an estimate of how many poor people live in a country, for instance.

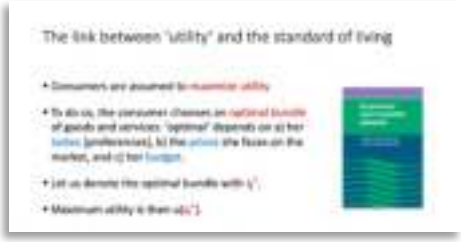
The next few slides introduce the views of a notable economist and philosopher, Amartya Sen. Sen’s position is very critical of the simplification that mainstream economics adopts in order to arrive at a measurement of living standards – which will be covered in the rest of the lecture. However, the quote from Sen’s book ‘Commodities and Capabilities’ is a good summary of the complexity of the concept of well-being, and conveys the need of delimiting the issue, if one wants to make this concept measurable.

This slide brings the audience back to the fundamental question of defining the standard of living, after having established that such a definition is a complex matter, not just intuitively (see the short discussion that the class just had about well-being), but also in the literature (see Sen’s thought). However, mainstream economics provides a way to **reduce** this **complexity**. It does so by introducing the concept of “utility”.

The next slides explain the basics of **utility theory**. If more details are needed, students can be encouraged to consult a manual of microeconomics (chapters on consumer choice). One of the most widely used is Varian (2010), while Deaton and Muellbauer (1980) or Varian (1992) are at a more advanced level.

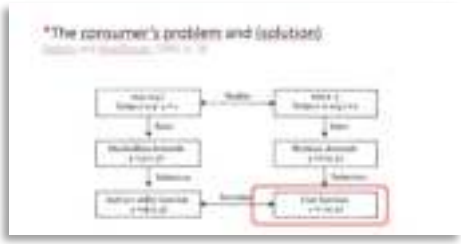


The answer that mainstream economists give to the complexities of defining well-being is based on the concept of utility, introduced in this slide. “Utility” is an abstract term, a catch-all, standing for whatever a person pursues. It is a useful tool for standard economic theory, which is concerned with individual choice.



Individuals (consumers) are assumed to derive utility from the consumption of goods and services, which means that utility depends on q (a vector of goods and services). In this theoretical setting, consumers are assumed to make their choices based on utility. How? They choose the particular combination of goods and services (also called a “bundle”) that gives them the highest possible utility, given the circumstances. These circumstances include the consumers’ individual tastes (also called preferences), the prices that they face on the market, and their budget (the resources they are able to spend on consumption). In other words, consumers are assumed to **maximize utility**. Consider a single consumer, and let us introduce a bit of notation: the optimal bundle (the one that the consumer ends up picking) is indicated by q*, and therefore the consumer’s maximum utility is u(q*).

We would like to know how much is u(q*), because that would be a measure of the consumer’s living standard. However, we cannot observe utility: it is just an abstract concept. What we can observe is the cost of q*, which can be indicated by x = q*× p (where p are the prices paid by the consumer). Welfare analysts follow Deaton and Zaidi (2002), a paper that shows how to calculate the cost of utility given the cost of the bundle.



Optional slides are denoted by a ‘star’ (*). Depending on the circumstance, we suggest that, especially the first times the course is thaught, these slides are omitted from the

presentation, unless we feel confident that it will not take to long to convey the message, and the audience is well equipped to absorb its content. This optional slide goes into more detail on one key point: the fact that the cost of the optimal bundle indicates the utility achieved by the consumer when she chooses a given bundle. The justification for this is the equivalence between the utility maximization and expenditure minimization problems, known as *duality*. A detailed discussion is in Deaton and Muellbauer (1980), p. 28. In what follows, we assume the Instructor has *not* shown to the class and commented on the primal and dual consumer problems in the starred slide above.

Deaton and Zaidi (2002) show that the utility associated to the optimal bundle can be approximated by (calculated as) the cost of the optimal bundle, divided by a price index, which transforms the expenditure into real terms, correcting for differences in purchasing power that arise across different consumers. Price indices can be computed according to the Paasche or Laspeyres formula: these details are not important for now, and they will be covered in the following lecture. Suffice it to say that x/P is called money-metric utility (MMU), and it is the quantifiable concept of living standard that modern economic theory offers. In practice, this concept corresponds to total consumption expenditure adjusted with a Paasche price index (the index will be explained in detail in lecture 2 – so the suggestion is not to waste time going into details). This is what equation (2.6) in Deaton and Zaidi (2002), possibly the single most important result underlying the way we measure consumption in poverty and inequality analysis, establishes. We do not expect students to get down to the details and the math required to derive this equation, but it is important to explain that this recommendation is not one of many available, but the only one consistent with modern economic theory. Other options, no matter how attractive, do not have this advantage.

In practice, Deaton and Zaidi's recommendation – to use total consumption expenditure divided by a Paasche price index to proxy living standards – answers the question posed at the beginning of the lecture: **what is the standard of living? x/P** is no longer abstract, or unobservable, like the concept of utility. Instead, it is a concrete and achievable measure of well-being, consistent with economic theory.

This optional slide suggests further readings on the foundations of living standards measurement to interested students.

Because we are now dealing more and more with the *practice* of measuring living standards, some clarifications are in order. This slide makes some important distinctions: "expenditure", "consumption", and "consumption expenditure" are different concepts. "**Expenditure**" is the nominal money outlay of the household: simply put, it is the total amount of money spent by the household over a given reference period. "**Consumption**" refers to the quantity of goods and services that the household enjoys, or uses, during the reference period. "**Consumption expenditure**" refers only to the expenditure that is directed toward consumption: in other words, the amount of money spent to buy goods and services that are actually used (as opposed to being stored, which qualifies as an investment) over the reference period. An example may help: suppose that over a given reference period, say one month, a household spends $30 to buy 30 kg of rice. Of these, 20 kg are actually eaten by the members of the family, while the remaining 10 kg are stored, and remain available for future needs (maybe they will be consumed, maybe they will be given away, or wasted: for now, we do not know). In this example, $30 is household expenditure; 20 kg is household consumption of rice; $20 is household consumption expenditure.

The next slides recap the take-home points from the first part of the lecture. There are a number of different approaches to measuring living standards, which were not all covered here; so far, the focus has been on one approach, which is put forward by mainstream economics. Economists seek to measure utility, which they approximate by consumption

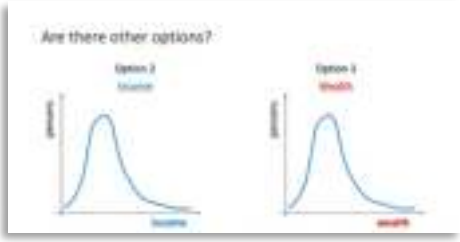
expenditure, adjusted for purchasing power. Granted, expenditure excludes potentially important contributors to utility, such as publicly provided goods, or leisure – anything that does not have a measurable price. It also measures a "simplified" concept of well-being which Amartya Sen critiques, as seen in the opening slides. However, despite these limitations, this approach is advantageous enough to be the preferred one for the World Bank's analysis of global poverty.

This optional slide mentions an important exception to the points just made: **multidimensional poverty measurement** is increasingly popular around the world. The cited reference is suggested to interested students.

Now is a good time for a break.

After having laid the theoretical foundations for the choice of consumption expenditure as a measure of living standards, we move to considering other potential candidates, some of which are very common in practice: the final part of the lecture deals with the pros and cons of these competing measures.

The choice of **consumption expenditure** adjusted for purchasing power follows directly from the theoretical framework of money-metric utility, which was just discussed.



Other potential candidates are: household **income**, that is, receipts, whether monetary or in kind (goods and services) that accrue to the household or to individual members of the household over a given period of time (labor income, capital income, public and private transfers received); and household **wealth** or net worth, that is, the value of savings, investments, real estate and cash, less any debts, possessed by the household or by its members at one point in time.

Should these other options be considered viable candidates for the purpose of measuring living standards? **Wealth** is easy to rule out, therefore it is examined first. While it seems logical to believe that wealth contributes to the standard of living, there are both conceptual and practical arguments against it. Conceptually, wealth can be defined as a stock of resources. While consumption expenditure is a flow, which means that it is defined over an interval of time (what was spent over a day, or a month, or a year), wealth is measured at one specific point in time (say, today at midnight), and may have accumulated in the past. Individuals may or may not choose to use their stock of wealth to generate consumption (and thus utility) in the present, which is what we care about. There are also practical reasons that discourage the use of wealth as an indicator of living standards: it is difficult to measure, and data on wealth are only rarely available.

Income, on the other hand, seems like a viable alternative: in fact, many countries use it.

An example of the difference between income and consumption expenditure clarifies that we are talking about two very different objects. The answer to the final question, "which of these two conclusions is correct?", is, in a way, ambiguous. If one accepts the conceptualization of standard of living that was covered in the first part of the lecture, then consumption expenditure is the correct measure, because it captures the value of use of commodities (and therefore, money-metric utility). However, the choice of income can be justified by a slightly different concept of standard of living, where the emphasis is on potential rather than actual consumption. This concept turns out to be particularly fitting when the main focus is inequality, rather than poverty.

The work of Sir Anthony B. Atkinson, a well-known scholar and expert in the field of inequality, is a testimony to this statement. This slide shows quotes from a recent book by Prof. Atkinson, where this alternative conceptual framework is briefly explained. This is just a quick hint, as the main frame of reference for poverty measurement, and for this course, remains money-metric utility; interested students can go to Atkinson's book for more information.

The next slides list some of the **advantages and disadvantages** of income and consumption expenditure as measures of living standards. We suggest to comment on each bullet rapidly, inviting students to think about how convincing each argument is given the context – we emphasize that context matters, that there is an on-going discussion among experts on the relevance of these arguments, and no easy way exist to reach an agreement. What is important is awareness of these issues, and ability to assess each of them in a specific context.

The **advantages of income** are the following: *(i)* in practice, household-level totals of both consumption expenditure and income are often computed as a sum of sub-components (expenditures in various items, various sources of income), and sources of income are fewer in number than consumption items: this makes it easier to gather information on income, at least in this regard; *(ii)* some sources of income are directly connected to the individuals earning them, which allows for analyses of individual living standards in a way that consumption expenditure does not; *(iii)* the final point reiterates what was covered earlier: income is consistent with a concept of well-being as potential command over resources, and this makes income a good candidate when analysts are interested in such an approach.

The **disadvantages of income** are the following: *(i)* income is frequently affected by short-term fluctuations, especially in rural areas: we are usually interested in a measure of living standards that is representative of a longer reference period, say a year (more on this later), and it is challenging to adjust for these fluctuations at the data collection stage, and at the analysis stage; *(ii)* under-reporting is a serious issue when it comes to collecting data on income, and exposes the measure to significant measurement error; *(iii)* some income components are difficult to observe: two cases that are especially relevant in poorer countries are that of revenues from informal labor, and that of valuing incomes from agricultural home production (which are in-kind).

The **advantages of consumption expenditure** are the following: *(i)* the choice of this indicator is grounded in utility theory (the framework that was developed at the beginning of the lecture), and this is a desirable feature for empirical analysis; *(ii)* consumption is typically easier to recall than income, which decreases measurement error; *(iii)* again, the

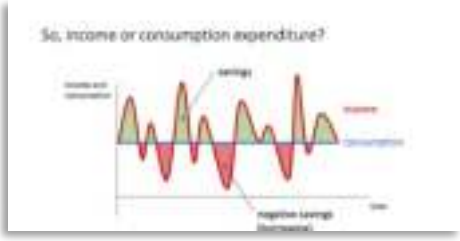
final point reiterates the consistency of consumption expenditure with a concept of well-being that coincides with the actual fruition of resources, which makes it a good candidate if that is the target of analysis.

The **disadvantages of consumption expenditure** are the following: *(i)* certain expenditures are directed toward items that do not enhance well-being (some economists call these items "bads", in contrast with the term "goods" that is used generally): fines are an example; *(ii)* families may make extraordinary expenditures, that are not representative of their usual living standards: weddings and other social functions are an example; *(iii)* some components of consumption expenditure, like the use-value of durable goods and housing, are difficult to measure and require estimation procedures (more on this in a coming lecture).

So, what is the balance of the pros and cons that have just been reviewed? As previously mentioned, the choice depends on the goals of the analysis and on the context. However, a final argument can be made regarding pros and cons of the two competing measures, and that concerns the reference period, which measured living standards are supposed to be representative of.

According to the simplest models of consumer choice presented in textbooks of economics, **time** does not exist. Individuals make their consumption choices just once, selecting the optimal bundle, and spending all of their available budget. In this extreme case, income and consumption expenditure are the same. However, in real life, time matters. In general, income and consumption expenditure can differ*:* income = consumption expenditure + savings (where savings may be negative, which is called dissaving, and occurs when there

is borrowing). We are interested in the living standards of households over a given period of time, typically **a year**, and we want our estimates to be representative of that period.



The implication of this discussion for our choice of income versus consumption expenditure is explained by this figure. If the reference period we are interested in is a year, then we must evaluate income and consumption expenditure (or even wealth, or a combination of all three) over that period of time. Empirically, consumption expenditure fluctuates less (in other words, it is **smoother** over time) in the short-run than income does – refer to Deaton and Zaidi (2002: 14) for a discussion of this topic. The graph helps intuitively grasp what "smoothing" is.

This slide wraps up the discussion on **income vs. consumption**: among the advantages of using consumption expenditure as an indicator of living standards, we must count its (comparative) stationarity over time: because of smoothing, observing consumption over a relatively short period – even a week or two – tells us a great deal more about annual (or even longer period) living standards than income can.

At this point of the presentation the audience is most likely tired… so the way we conclude is by showing them something concrete, and hopefully interesting. This bunch of slides provides some examples of the choices made by countries around the world when it comes to measuring poverty and inequality, to provide a sense of what the common practice is, and how it varies. This overview is meant to give students some practical examples, and to ground the concepts covered in the lecture into practice.

The final slide is a recap of the whole lecture: it is a good idea to spend a few minutes on it, to make sure that the main bring-home conclusions are clear. The first point is especially important, and justifies the time and effort spent on the theoretical issues that were the focus of the lecture: the concepts and vocabulary just introduced are of paramount importance for data providers, not just for analysts, because data quality depends on **relevance** for a specific research objective. If data providers and analysts do not share the same conceptual tools, data will never match user needs and expectations, no matter the amount of money and effort spent in collecting them. The remaining points summarize the key messages of the lecture.

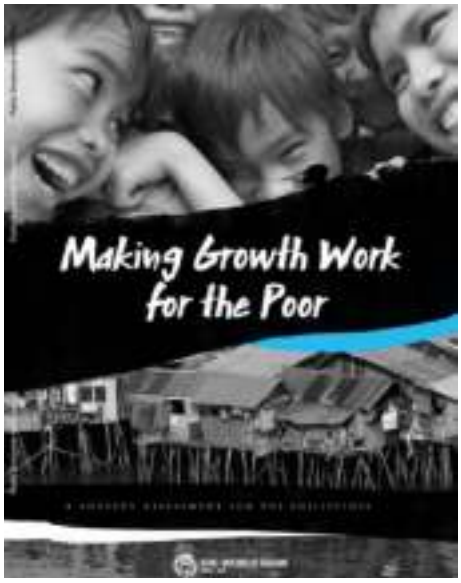
Homework

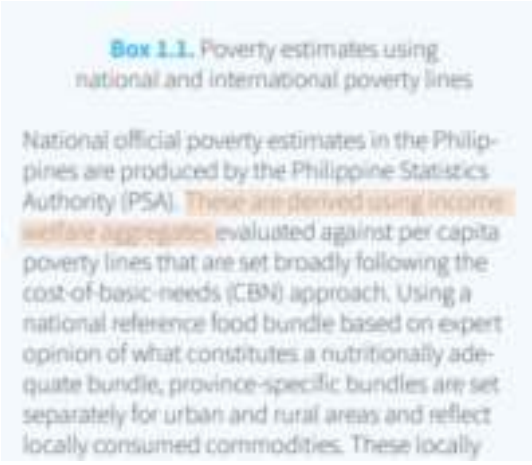
Exercise 1 – Engaging with the literature

This exercise is conceived as an invitation to get involved with relevant pieces of literature that are related to the topics covered in the lecture. The emphasis here is not on the essay itself, which should be brief and synthetic, but on the student's understanding of the key conclusions of the papers, and how they relate to the lecture. Obviously, there is not a specific and unique solution for this exercise: the Instructor should evaluate each short essay by keeping in mind the intent that was just described. The same holds for all "Engaging with the literature" exercises in the course.

Exercise 2 – Income or consumption?

This exercise gives students an opportunity to see how the theoretical principles covered in the lecture operate in practice, by exploring actual Poverty Assessment reports and the analytical choices they make. Students may find the cross-country variation in the choice of income or consumption interesting, and may also face instances in which the indicator of choice is not entirely clear – which is a valuable lesson in transparency and accountability. An example of what students may find is from the latest for the Philippines (2018), "Making Growth Work for the Poor: A Poverty Assessment for the Philippines":





Box 1.1. Poverty estimates using national and international poverty lines

National official poverty estimates in the Philippines are produced by the Philippine Statistics Authority (PSA). These are derived using income/welfare aggregates evaluated against per capita poverty lines that are set broadly following the cost-of-basic-needs (CBN) approach. Using a national reference food bundle based on expert opinion of what constitutes a nutritionally adequate bundle, province-specific bundles are set separately for urban and rural areas and reflect locally consumed commodities. These locally

Exercise 3 – Multidimensionality of well-being

The goal of this exercise is to show that drawing conclusions on a country's living standards based on multiple indicators of well-being, while attractive (and useful in some contexts), is challenging, as different indicators do not normally agree with each other. Once again, the emphasis here is not on the essay itself, but on students' ability to engage critically with some of the concepts covered in the lecture.

Lecture 2
The consumption aggregate

Learning objectives

The goal of this lecture is to complete the conceptual framework underlying the construction of a measure of living standards. Building on the conclusion reached by Lecture 1 – preference should be given to a consumption-based measure of living standards – this lecture provides a working definition of consumption expenditure, and discusses the adjustments that analysts should apply to it, in order to construct a *living standard indicator*.

Suggested preparation

The Instructor is assumed to be familiar with the material contained in sections 4 and 5 of Deaton and Zaidi (2002).

Time allocation

Roadmap and 'which expenditures?'	**15 min**
Adjusting for household size and composition	**30 min**
Adjusting for purchasing power	
Terminology	10 min
Break	**15 min**
Inflation example, price indices	45 min
Lessons learned	**5 min**

Annotated lecture

The lecture opens with the agenda for the day.

This slide provides the roadmap for the whole lecture. Our objective, constructing a living standard indicator on the basis of survey data, consistently with the theoretical framework covered in Lecture 1, is going to be reached by combining four elements, as shown here. This lecture will be discussing each of the four elements.

The next few slides discuss the numerator of the living standard indicator: nominal household consumption expenditure. The takeaway for this section is that this step does not simply imply adding up all household expenditures, precisely because what matters is *consumption*, rather than *expenditure* per se. The discussion of this topic is intentionally kept short, as it relates more to analysis than to data collection. More space will be given to adjustments (the denominator of the living standard indicator), which have more bearing on survey design choices.

This slide signals that the numerator of the living standard indicator has been dealt with: it is time to move on to the next component, that is, the adjustment for household size and composition.

Some introductory slides explain why we should concern ourselves with differences in household size (including the possibility of accounting for economies of scale) and household composition.

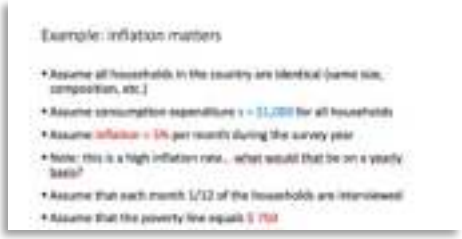
The following batch of slides explains the most common approaches used in practice to adjust for economies of scale (using the coefficient ‘alpha’ to modify household size) and household composition (using equivalence scales), and explains in which cases these adjustments are recommended.

Then, some evidence on what countries around the world actually implement is presented. Different regions of the world adopt different strategies, but it is shown that expressing the living standard indicator *at least* on a per capita basis is routine.

Again, this slide signal that the lecture is now moving on to the last topic: the adjustment for differences in purchasing power, which encompasses elements 3 and 4.

These slides provide some basic terminology: nominal vs. real, consumer price indices (CPI), temporal and spatial deflation. The way CPIs should be interpreted is also clarified.

Now is a good time for a break.



These slides present a hypothetical example, that demonstrates why adjusting for differences in purchasing power is important. The example focuses on temporal price differences, i.e. inflation, but the same reasoning holds for spatial price differences. The upshot is that using nominal or real expenditure produces different poverty estimates, and that the second measure best represents actual living standards. A real example from Kenya concludes the discussion, showing the relevance of this issue in practice.

The final batch of slides discusses different price indices, namely Laspeyres and Paasche. Section 4 of Deaton and Zaidi (2002) is the main reference for this topic. The goal is for students to understand that there are several ways to adjust for price differences, and that they are not equivalent. When the focus is welfare measurement, the preferred price index is Paasche. Again, the discussion wraps up by showing a real example of spatial differences in prices.

The final slides summarize the topics discussed in the lecture, with a focus on their implications for survey data collection.

Homework

Exercises 1 and 2 – Engaging with the literature

See exercise 1, Lecture 1.

Lecture 3
Understanding household surveys

Learning objectives

The goal of this lecture is to provide an overview of the many different survey instruments developed and implemented by statistical institutions around the world. The material covered in the lecture puts household consumption and expenditure data – the main focus of the course – into a broader context.

Suggested preparation

The paper by Grosh and Glewwe (1998) provides background information on the Living Standard Measurement Study and the surveys developed under its umbrella. Familiarity with it is assumed.

Time allocation

Overview of statistical instruments	**20 min**
Overview of household surveys	
Introduction and terminology	30 min
Break	**15 min**
Taxonomy: common surveys, quick HCES, large HCES	50 min
Lessons learned	**5 min**

Annotated lecture

The set opens with a roadmap: after concepts, which were covered in the previous lectures, it is now time to move to data. There are many different ways for statistical institutions to collect data, and this lecture provides broad classifications and terminology for different data collection efforts.

Evidence of an increasing demand for data – regarding all aspects of society – is used to justify the multitude of different statistical instruments currently in use.

A classification based on two dimensions – representativeness and objectivity – can be a useful tool to evaluate these different instruments. Notable examples of information gathering are placed on a diagram defined by these two dimensions. Note that a *windscreen survey* is not a survey at all, but rather an informal, subjective assessment of one's surroundings, done from one's seat in a car; and a *poverty participatory survey* is similar to a focus group, ran with the aim of understanding poverty determinants by incorporating the perspectives of the poor themselves.

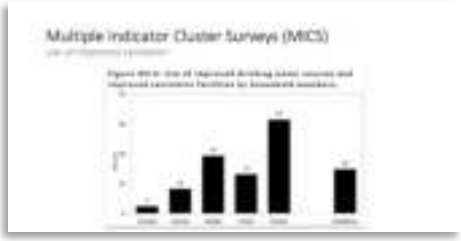
Representativeness and objectivity are desirable properties, and they help restrict our focus to statistical instruments that possess these qualities: household surveys.

The topic is introduced with some historical background and some terminology (for 'household' and 'probabilistic sample'). Once the basics are covered, a slide introduces a

classification of the main types of household surveys. On the one hand, we have ‘common’ surveys – common in the sense that they are virtually ubiquitous, frequently mentioned, essentially a staple for most statistical offices around the world. On the other hand, we have HCES: not necessarily less ‘common’ than other surveys, what characterizes them is their focus on measuring consumption and expenditure; they can be divided into ‘quick’ and ‘large’ surveys. The following slides go into each of these groupings.

Now is a good time for a break.

...



The group of slides on common surveys is structured as follows: each survey type (LFS, DHS, MICS) is introduced with a general description; a recent example from a country is presented, to show what the survey might look like in practice; a final slide shows an indicator or statistic that is usually computed from that type of survey. This section can be covered rather quickly, as students may already be familiar with some of the material.

...

The discussion then moves to HCES. ‘Quick’ surveys are described synthetically, as they are complementary to ‘large’ surveys. Interested students are directed to the suggested references for more details. Optional (starred) slides discuss an example from Indonesia, and the Nigeria CWIQ 2006.

Finally, the lecture zooms into 'large' HCES. This part should be emphasized over the others, as it is more important for what will follow in the rest of the course. In particular, HBS and LSMS are two main approaches to conducting HCES, they are both routinely used for measuring living standards, poverty and inequality, and the differences and similarities among them should be adequately discussed.

The bring-home conclusion is that welfare analysts need household budget data (e.g. HBS, and LSMS type of surveys). By the end of the lecture, the characteristics of this type of surveys, and the general context in which they are situated, should be clear to students.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Household?

This exercise gives students an opportunity to see how the theoretical principles covered in the lecture operate in practice, by exploring actual survey documentation material and the definitions underlying the design of questionnaires. An example of what students may find is from the enumerator manual of Malawi's Integrated Household Panel Survey 2010-2013:

"A household may be either a person living alone or a group of people, either **related or unrelated**, who **live together** as a single unit in the sense that they **have common housekeeping arrangements** (that is, share or are supported by a common budget). A standard definition of a household is: "**a group of people who live together, pool their money, and eat at least one meal together each day**". It is possible that individuals who are not members of the household may be residing with the household at the time of the survey. In most cases, but not all, someone who does not live with the household during the survey period is not a current member of the household."

The students' findings can be used as a basis for discussion of the differences, or similarities, of such a fundamental definition across countries.

Exercise 3 - Household surveys

This exercise encourages students to identify the main features that characterize a survey, and pushes them to read and get familiar with technical documentation. Again, there is no single right answer for this exercise: the students' findings can be used to start a discussion on the topic of harmonization. A good solution might look like this:

country	survey	year	Sample size	Sampling method
Tanzania	DHS	2015-2016	13,376 hh	Stratified sample selected in two stages. Probability proportional-to-size selection at the first stage of sampling
Tanzania	NPS	2014-2015	3,265 hh	multi-stage cluster sample design.
Ethiopia	DHS	2016	18,008 hh	Stratified sample selected in two stages. Probability proportional-to-size selection at the first stage of sampling
Ethiopia	ESS	2015-2016	5,469 hh	Two stage sampling. Probability proportional-to-size selection (of EAs) at the first stage of sampling
Zambia	LCMS	2015	12,260 hh	two-stage stratified cluster sample design. Probability Proportional to Estimated Size (PPES) at the first stage of sampling
Zambia	LFS	2014	11,520 hh	disproportionate allocation to strata

Exercise 4 - Sherlock Holmes

The goal of this exercise is to strengthen the students' awareness that any aggregated indicator originates from survey data, and that different surveys are used for different purposes (typically, labor force and poverty statistics are computed on the basis of Labor Force Surveys and Household Budget Surveys, respectively).

Exercise 5 - The Gemini Project

The goal of this exercise is to acquaint students with the Gemini Project, an important initiative launched to help redesign the US Consumer Expenditure Survey, addressing issues of measurement error and respondent burden. The Project has spurred considerable research effort, and many of the papers published under this umbrella contain findings that are relevant for other countries.

Exercise 6 – Historical Household Budget Surveys

This exercise is similar to exercise 3, but this time the focus is on considering the issue of harmonization of survey methods over time.

Lecture 4
Principles of questionnaire design

Learning objectives

The goal of the lecture is to review some general principles of questionnaire design, as they apply to the design of surveys to measure consumption. Principles include both general planning recommendations, and specific tips on formulating questions, applicable to any module of the survey.

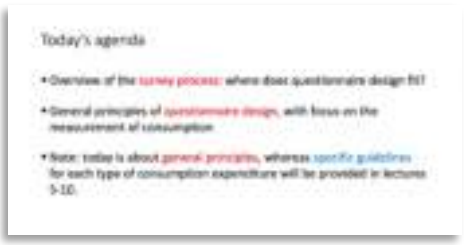
Suggested preparation

Chapters 2, 3 and 5 of Grosh and Glewwe (2000) are a classic reference for the topics reviewed in this lecture, and are especially useful given their focus on LSMS surveys. The more recent Glewwe (2005) revisits some of the same material, in a more synthetic fashion. Familiarity with both works is assumed. Reviewing chapter 2 of Biemer and Lyberg (2003), and chapter 3 of Iarossi (2006) is also advised.

Time allocation

The survey process	**20 min**
Principles of questionnaire design	
Choice of topics and respondents	40 min
Break	**15 min**
Formulation of questions, order of questions, field testing	40 min
Lessons learned	5 min

Annotated lecture



The set opens with the agenda for the day. We remind students that the lecture will focus on **general principles** underlying the questionnaire design stage, and that guidelines that apply to the design of individual modules will be discussed in upcoming lectures. We also

invite students to engage with the recommended readings, which provide the material covered in this lecture.

First, it is useful to put questionnaire design into **context**, among the different steps that make up the survey process. The design of the questionnaire is informed by research objectives, which in turn are the goal of data analysis: the idea that the last phase of the whole process – the actual use of the data – defines priorities at the very beginning of it is represented by the arrow that circles back from step 5 to step 1 in the figure.

The papers cited in these slides, Joliffe (2001) and Backiny-Yetna et al. (2017), are here as an example of the fact that questionnaire design makes a great deal of difference to final estimates. Both papers will be discussed further in the next lecture, so there is no need to spend much time on this part. Suffice it to say that these are two examples of research comparing questionnaires that differ only in seemingly small and irrelevant features, and conclude that these features are responsible for large differences in the estimates of interest. The takeaway is: **details matter**.

This diagram shows the steps that a questionnaire design team follows when constructing a questionnaire. We will discuss them one by one.

First, a prerequisite for the whole process: the **definition of research objectives**. The Instructor should review the references cited in grey for more details.

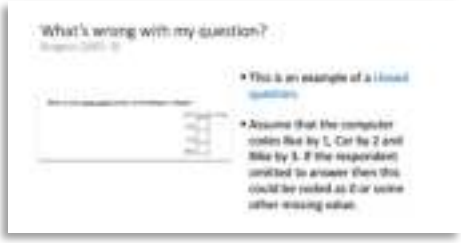
After having defined the research objectives addressed by the survey, it is good practice to proceed by choosing broad **topics** to be included in the questionnaire. These two slides deliver general advice – adopting a **module approach** – and cite the LSMS survey template as an example.

Next, who in the household should be addressed as a **respondent**? It is recommended to select whomever is **most knowledgeable** on the question or module at hand. Deaton and Grosh (2000) point out that for some consumption items, it is unrealistic to expect a single person to have an accurate picture of each household member's own expenditures. In some cases, individual modules are indeed preferred (some examples will be cited in the next lectures: one is food away from home) but the benefits of implementing such an approach should be weighted against the increased costs with respect to a 'proxy respondent' module.

Now is a good time for a break.

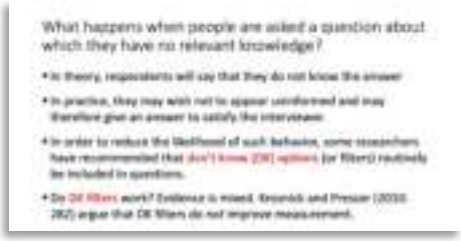
What to ask, and how? The next lectures will address the 'what'; for now, a few words of advice on the 'how' (formulation of questions).

In terms of **question wording**, a useful rule to remember is the so-called 'BOSS principle': questions should be brief, objective, simple, and specific. A few slides explain each criterion with an example: students can be encouraged to try and give their own answers before we discuss 'possible fixes'.



The choice between an open or closed question is about **question type**. Closed questions with pre-coded answers are most common in expenditure surveys: there are a few basic recommendations that help to design them correctly. The example reproduced in the 'bus, car, bike' question highlights the importance of well-designed response codes. The problem with this questions is that, although codes are mutually exclusive, they are not exhaustive: there is no option for 'none', which leads to confusion between 'none' and missing.

Another recommendation on pre-coded question is to allow enough flexibility to accommodate the heterogeneity of respondents and reduce the cognitive burden of answering an unfamiliar question.



This slide touches on 'DK filters'. The Instructor is directed to the reference provided for more details.

We remind students how much ground was covered, and where we are in the survey process.

The order in which questions are asked (the question 'flow') is also important. One of the main topics to have been discussed by the literature is that of sensitive questions, and how to ask them: the paper by Barton (1958) is a funny way to convey the issue. A common solution in modern surveys is the **unfolding brackets** approach, which is often used for questions on income.

Finally, a few tips about pre-testing the questionnaire. The fact that no questionnaire is ever perfect, and that arriving to a good design is likely to be an incremental process, leads to

some considerations about the benefits and pitfalls of accuracy (continually improving the design of the questionnaire) and those of comparability (adopting the same design as previous waves. These considerations are echoed in the final slide (lessons learned).

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Question wording

Questions 1 and 3 are too subjective: they will not necessarily yield the present value of the house (according to the market), but how much it is worth according to the owner. Question 2 is worded in a much too technical way, with excessive jargon. Question 4 is just right: it is easy to understand, but at the same time grasps the right concept (what people would be *willing* to pay for the house). In fact, this question is often found in the housing section of many HCES. Question 5 is inaccurate: it does not ask for the net present value of the house, but for its original purchase price.

Lecture 5
Measuring food consumption: the foundations

Learning objectives

This lecture, together with lectures 6 and 7, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 5 focuses on the concepts of *acquisition* and *consumption*, and on determining the mode of data collection (recall vs. diary).

Suggested preparation

Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The instructor is assumed to be familiar with both. This lecture also summarizes the evidence from a few important studies (Beegle et al. 2012, de Weerdt et al. 2016, Backiny-Yetna et al. 2017), which the instructor is encouraged to review.

Time allocation

Introductory slides	**5 min**
Acquisition vs. consumption	
Concepts	20 min
Examples from questionnaires	15 min
Recommendations	10 min
Break	**15 min**
Recall vs. diary and length of reference period	
Concepts	20 min
Evidence from Tanzania and Niger	20 min
Recommendations	10 min
Lessons learned	**5 min**

Annotated lecture

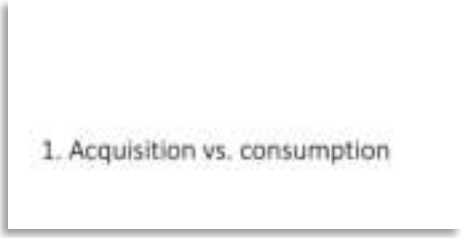
The set opens with a brief introduction, which has the goal of re-acquainting the students to the main topic of the course – measuring consumption – after the brief detour taken during Lecture 4 on general principles of questionnaire design.

We begin by saying that today's lecture is the first of a group of three lectures (5, 6, and 7) focused on **measuring food consumption**. What justifies such a strong focus on this topic? First and foremost, food consumption expenditure is a key component of any measure of living standards (lectures 1 and 2), and therefore, it is needed for poverty measurement. But there are several other important research topics that justify an interest in measuring food consumption with household surveys. We may 'wake up' the audience by soliciting answers (2 mins discussion might help). Either way, it should be clear that accurately measuring food consumption is important, for different types of data users and researchers.

This and the following lectures go into the specifics of questionnaire design, and in particular, of the design of the food module. The main references for this group of lectures are the FAO/World Bank Guidelines on food data collection (left), and a recent paper by Smith, Dupriez and Troubat (right). The former illustrates best practices, while the latter is more focused on what countries actually do.

This slide is an outline of what is to come: the topics covered are the main 'challenges' that a questionnaire design team will have to face when implementing the food module of a household consumption and expenditure survey. The red rectangle indicates that Lecture 5

will deal with the first two topics (acquisition vs. consumption, and recall vs. diary), while the remaining issues will be tackled in Lectures 6 and 7.



Acquisition vs. consumption is our first topic.

The distinction between acquisition and consumption is a conceptual one, regarding *what* exactly is captured by the food module. The first of the slides pictured above gives rigorous *definitions* of acquisition and consumption (and adds some terminology on *modes* of acquisition); the second slide, "What to do with the chicken?", clarifies these concepts by means of an *example*. For more details, you can check Section 2 (p. 44) of Conforti, Grünberger and Troubat (2017), which elaborates further.

Does the distinction between acquisition and consumption really matter? Acquisition and consumption, as we have just seen, are different. Neither is intrinsically superior: each may be useful for different analytical purposes (the slide lists only a few examples). What is important is that the choice of what to measure be consistent with these objectives, and that the measure of interest be clear and unambiguous for both interviewers and respondents.

We now transition **from concepts to practice**. What do household surveys around the world record: acquisition or consumption? The assessment by Smith et al. (2014) shows that both concepts are routinely measured in practice. However, poor questionnaire design is common: it will soon be clear what this means, specifically (slides on 'questionnaire design issues' are coming up). Conforti, Grünberger and Troubat (2017) summarize the main approaches to data collection found in practice (neither of which, again, is intrinsically superior): check page 44 of their paper for more details.

The idea of flaws in questionnaire design is introduced by some examples that students are encouraged to interact with. We suggest to ask students to comment on a few questionnaires, by answering three questions. Questions 1 and 2 (see slide above) are purely descriptive: would the data collected through each questionnaire be suitable to measure food consumption, acquisition, purchases, or unit values? These questions check whether the concepts covered in the previous slides are clear to students. Question 3 encourages students to think about possible flaws in the way the questionnaires are designed. More of these examples are used as homework for this lecture.

This page is from **Burundi** QUIBB 2006. It is *not* possible to fully capture consumption or acquisition with this questionnaire. If the objective is to measure *consumption*, after the filter question (M11) data are collected on food acquired (purchased, harvested) rather than food consumed. When this occurs, quantities and expenditures on food acquired include those entering into the households' stocks – not the household pantry for immediate consumption – and are systematic overestimates of food consumed from home production. If the objective is to measure *acquisition*, then question in M11 rules out any following question for food that were acquired but not consumed in the reference period. These are flaws in questionnaire design, because they lead to incomplete enumeration of both acquisition and consumption.

On the other hand, *unit values* can be computed based on this module: it would be sufficient to divide M13 by M14 (provided local units can be converted into standard units).

This example is from the **Madagascar** Enquete Permanent Menages (EPM) 2005. The questionnaire captures *acquisition*, rather than consumption, although in this page we cannot see any question on own-produced foods, which would therefore be missing from the total amount and value of acquisitions.

Unit values can be obtained as question 6 divided by question 7.

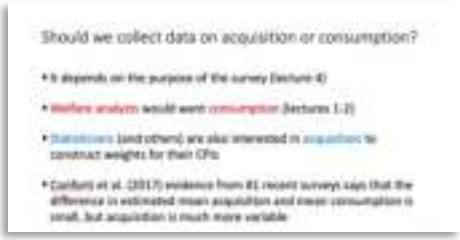
A problem with the design of this questionnaire, besides the apparent lack of questions on own-production, is that it is not clear if the "average per month" in questions 6 and 7 refers to the average of the past 12 months, or the average over those months in which consumption actually took place. A better approach would have been to clarify the ambiguity, or even to get rid of the "usual month" approach (more on this later).

This page is from the **Iraq** Household Socio Economic Survey 2012. The questionnaire only records food *acquisition* and not consumption. However, all modes of acquisition are recorded, and there are no apparent flaws in the design of the questionnaire. *Unit values* could be computed as question 1203 divided by 1202 (provided that non-standard units are convertible into kg).

This page is from the **Lebanon** Household Budget Survey. This questionnaire partially records food *acquisition*: it focuses on *purchases* only, and does not record food acquired through other sources. On the other hand, the questionnaire captures food *consumption* in full: question 7 allows to record all sources of food consumed. There are no flagrant mistakes in questionnaire design, assuming that the goals of the survey did not include

measurement of total food acquisition. Again, *unit values* may be computed both from purchases and consumption, by computing the ratio between reported value and quantity.

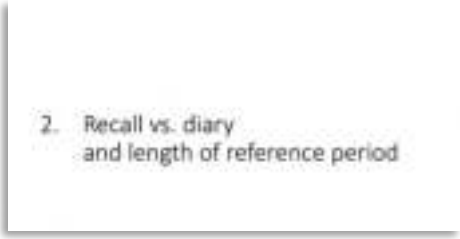
This slide summarizes the most common issues found in household survey questionnaires in developing countries, as listed by Smith et al. (2014). Some of these were exemplified by the questionnaires discussed with students. These should be seen as a catalogue of the most common mistakes to be avoided when designing the food module.



Provided that mistakes and incomplete enumeration are avoided, one question remains: **is it best to record acquisition, or consumption?** The answer depends on the goals of the survey. Based on the conceptual framework covered in Lectures 1 and 2, we could argue that welfare analysts prefer food *consumption*. However, recording food *acquisition* is important for other purposes, the most common of which is probably the computation of consumer price index (CPI) weights. The evidence presented by Conforti et al. (2017) is reassuring: the two measures are not too distant, so that one may be used in place of the other, in case of necessity. A common approach in practice, as we have seen, is to record both.

A set of recommendations closes this topic. Recommendations are adapted from section 3.3 of the FAO/World Bank Guidelines (2018): further details are found there.

Now is a good time for a break.



Recall vs. diary, and the determination of the length of the reference period is our next topic for the lecture.

First, the slides give definitions for diary and recall, recall period and reference period. Two examples are used to substantiate the difference between diary and recall.

Which of the two approaches guarantees the collection of quality data? The answer is not easy: both diary and recall surveys have the potential to generate measurement error. Empirical evidence helps to evaluate the risks implied by each of the two methods.

The most problematic aspect of using recall methods is related to memory and its failures: respondents' inability to remember correctly leads to biases in data collected through recall surveys. Depending on the length of the recall period, different types of bias may arise. *Memory decay* and *telescoping* are related to long and short recall periods, respectively. For a detailed explanation of these biases, see the discussion at pages 110 and 111 of Grosh and Glewwe (2000) (the 'Consumption' chapter by Deaton and Grosh).

Another issue that emerges when using recall surveys, and that is related to the length of the recall period, is that of the variability of reported consumption. Shorter recall periods record 'episodic' consumption (consumption that takes place on a given day or week) which may be randomly higher or lower than usual. These anomalies are smoothed out by longer recall periods. As a result, data recorded with short recall periods are *more variable* than those recorded with longer ones. This is a problem if data users are interested in the 'tails' of the distribution of consumption – households with very low calorie intakes, for instance – because the likelihood of extremely high and extremely low will be overestimated by surveys using short recall periods. Crossley, Fiedler and Mwangi (2016) discuss this issue in rather technical terms on page 19 of their IFPRI Discussion Paper.

Using diary surveys to collect data on food consumption may seem like a solution to the memory biases that characterize recall surveys: with diaries, respondents note down their consumption as soon as it happens. In practice, however, the diary approach introduces other problems, mainly linked to respondent burden and fatigue, and to the high implementation costs needed to minimize them.

Some authors have suggested alternative methods, a ‘third way’ besides diary and recall. However, empirical evidence has failed to support these alternatives.

Now that the trade-offs of the diary and recall approaches have been covered, one may ask:

The following slides summarize results from a few notable experimental studies that tested the effects of different questionnaire designs on estimated food expenditure and other outcomes.

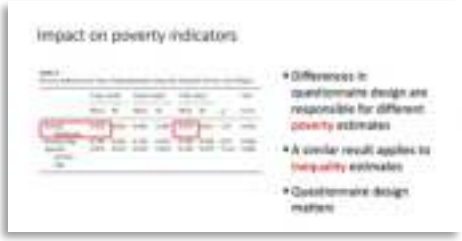
…

First is Beegle et al. (2012) with an experimental study conducted in Tanzania. The instructor is encouraged to read the paper: the slides offer a quick summary of the main findings, as they relate to the topics of Lecture 5. The study implemented eight different questionnaire designs, assigned them randomly to 4,000 households, and then compared average food consumption expenditure resulting from each of the questionnaires. One design – a personal diary with frequent visits – is considered the most accurate way of collecting data on food consumption, and is therefore used as a benchmark, against which all other questionnaire designs are stacked up.

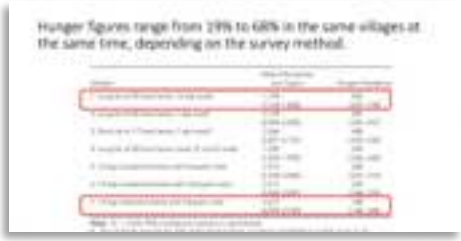
This slide shows the comparison of diary and recall questionnaires, by using a type of chart that is going to appear several more times in the lecture. Each design is compared to the benchmark (personal diary, frequent visits). The chart plots the coefficients of a regression of log consumption on dummies indicating assignment to questionnaire designs. Asterisks indicate significance: *** significant at 1%, ** at 5% level. No asterisks indicate no statistical significance (at traditional significance levels). Interpretation of coefficients is as follows (take the first line as an example): average food consumption expenditure obtained with a household diary with infrequent visits is 13.6% lower than the value

obtained with a personal diary with frequent visits (the benchmark design). Now, if we compare the coefficients for the two different *diary* designs, to those for the five different *recall* designs, we note that results are mixed. All numbers are negative, meaning that all alternative designs return a lower average food expenditure than the benchmark; but *how much* lower seems to depend on characteristics of the design *other than* the choice of diary vs. recall.

Other selected results of the study concern implementation costs for different designs (diaries give better results when closely supervised, at least for less-educated households; however, they are also much more expensive to implement than recall modules), and the length of the recall period (comparing the final two designs in the chart, whose only difference is the length of recall, shows that the 7-day recall comes much closer to the benchmark than the 14-day recall, likely because of memory decay).



Backiny-Yetna et al. (2017) is another experimental study, conducted in **Niger**. In this case, the comparison between diary and recall gives somewhat unexpected results: a 7-day recall module returns higher estimated food expenditure than a 7-day diary. This is reflected by poverty estimates (the poverty headcount is higher according to the diary module). Respondent burden is likely to be responsible for the low results obtained with the diary, which suggests that supervision matters.



Finally, De Weerdt et al. (2016) use the same data as Beegle et al. (2012), but compare results on nutrition. In this case, recall questionnaires yield lower estimated calorie intakes than diary designs.

To close the discussion of empirical evidence, we once again turn to the review of the current practice by Smith et al. (2014), to see which approaches are actually implemented in practice. In this graph, diaries are interpreted as surveys using a recall period of 1 day. Note that using different recall periods for different foods ('other' in the graph) is common.

Recommendations are adapted from Section 3.1 of FAO and World Bank (2018).

The final slide summarizes the main takeaways of Lecture 5: it is important to take some time for this final recap, as it provides an opportunity to cement the students' understanding of topics covered, and take any final questions.

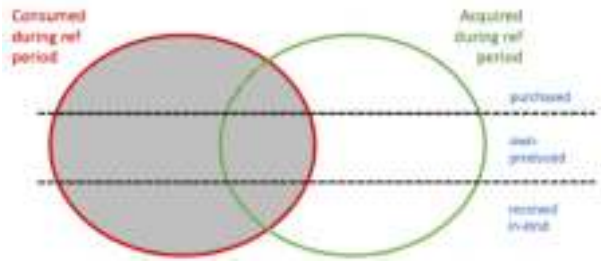
Homework

Exercise 1 – Engaging with the literature

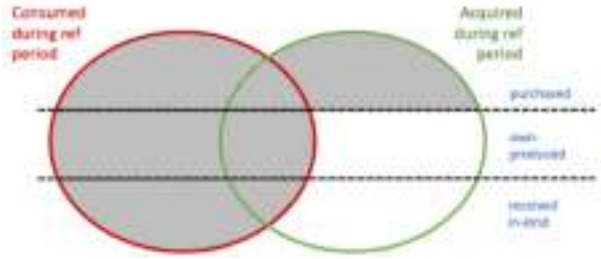
See exercise 1, Lecture 1.

Exercise 2 – Acquisition vs. consumption

Example 1 allows to estimate the total value of consumption. All three modes of acquisition are accounted for. The corresponding diagram is pictured below.



Example 2 also allows to estimate the total value of consumption. However, question C02 asks for the *total* value of purchases during the last 7 days (not just those going towards consumption): this makes it possible to estimate at least part of the value of acquisitions, precisely the portion acquired through purchases. The corresponding diagram is pictured below.



Lecture 6
Measuring food consumption: questionnaire design

Learning objectives

This lecture, together with lectures 5 and 7, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 6 focuses on the definition of the list of food items for which information is collected, on why and how to record meal participation, and on the timing of visits to interviewed households.

Suggested preparation

Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The Instructor is assumed to be familiar with both. This lecture also summarizes the evidence from several studies (Jolliffe 2001, Beegle et al. 2012, Pradhan 2009, Gibson and Rozelle 2002, Troubat and Grunberger 2017), which the Instructor is encouraged to review.

Time allocation

Introductory slides	**5 min**
List of food items	**40 min**
Meal participation	
Concepts: what partaking is and why it matters	10 min
Break	**15 min**
Evidence and examples, recommendations	20 min
Timing of visits	**25 min**
Lessons learned	**5 min**

Annotated lecture

After a brief introduction, aimed at reminding students the general framework in which the current lecture is inscribed and the outline for the day, we dive into the first topic: determining the list of items for the food module.

The goal of these first slides is to define what is meant by '**food list**', and to convey the fact that defining it from scratch is a difficult, non-trivial, and important task. The decisions that go into the definition of the food list are sometimes just summarized as "determining the length of the list".

It helps to think about the 'length' of the food list as resulting from two separate choices: the level of **comprehensiveness** of the list, and its **specificity**.

Comprehensiveness is, in a way, uncontroversial: the more comprehensive the list, the better. For this reason, the point is discussed quickly. Smith et al. (2014) set useful **criteria** that help to evaluate the level of comprehensiveness of the food list, and their assessment shows that most questionnaires perform well in this dimension.

Unlike comprehensiveness, **specificity** – the level of detail, or disaggregation, of the food list – has pros and cons. According to the literature, both too specific and too aggregated food lists can lead to measurement error. The studies cited make the point that specificity matters, in that it greatly affects final estimates.

The final **recommendations** give general guidelines to construct the food list. Section 3.6 of the FAO/World Bank Guidelines can be checked for further details.

Next, the lecture tackles the issue of adjusting for **meal partakers**. The goal of the first slides is to give a definition of partakers, and to motivate our focus on the issue, by conveying the message that knowing the exact *denominator* of per capita consumption (*how many* people are sharing the total?) is fundamental to obtaining an accurate estimate of individual welfare (be it consumption, expenditure, or nutrition).

Now is a good time for a break.

Empirical evidence is brought in to show that **adjusting for partakers matters** in practice. With the adjustment, estimated consumption decreases among the rich, and increases among the poor. This is a key reason why the adjustment is needed, given that estimating the *distribution* of consumption (not just the average) is one of the main uses for household consumption and expenditure survey. In contrast to the importance of collecting information on partakers, Fiedler and Mwangi (2018) show that most countries fail to collect data on this aspect, and those that do have very heterogeneous approaches.

These slides show some miscellaneous examples, in order of increasing complexity. For the time being, there is no need to comment on which approach is expected to work better: the examples are a way to give more substance to the idea of collecting data on partakers, and also to show how questionnaires can vary widely in practice.

These slides give more structure to the heterogeneity of approaches exemplified by the questionnaires shown. Smith et al. (2014) summarize the different pieces of information collected by questionnaires in developing countries; the FAO/World Bank guidelines reduce these different approaches to two main ways of adjusting for partakers – 'food consumer' and 'meal partakers'.

Next is the issue of determining the **timing of visits** to interviewed households. The choice matters because of fluctuations of consumption over time (seasonality, but also within-month or within-week variation).

Some evidence is presented on the **magnitude** of seasonal (and higher-frequency) variation in consumption: the main point to be made is that, because consumption is found to

fluctuate significantly within the survey year in most countries, the timing of interviews matters for final survey estimates.

This point is further stressed by this summary of the negative consequences of failing to adjust for seasonality.

The next slides discuss the best way to account for seasonality. The two most common approaches are introduced, and their frequency in practice is evaluated – together with the frequency of surveys that *do not* account for seasonality at all.

Recommendations prioritize the 'survey subsets' approach, although 'repeated visits' are also acceptable.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Food module, international comparisons

This exercise gives students an opportunity to see a practical application of the concepts discussed in the lecture, and become familiar with questionnaires used in different countries. The goal of the exercise is to assess the students' grasp of the guidelines on the

length of the food list, and see whether they can critique questionnaires based on the recommendations learned during the lecture.

Exercise 3 – Meal participation

Example 1 (Somaliland, 2012) collects information on individuals *usually* taking part in the household's meals. Data collected through this questionnaire would only allow to compute per capita intakes using the 'food consumer' approach (estimating the number of people usually partaking during the reference period).

Example 2 (Namibia, 2016) collects daily information on who takes part in each meal (breakfast, lunch, dinner). Data collected through this questionnaire would allow to compute per capita intakes using the 'meal partakers' approach (accounting precisely for the number of meals taken by household and non-household members on each day).

Lecture 7
Food away from home and the use of non-standard units for measuring food consumption

Learning objectives

This lecture, together with lectures 5 and 6, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 7 focuses on measuring the consumption of food prepared away from home, and on using non-standard measurement units.

Suggested preparation

Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The Instructor is assumed to be familiar with both. In addition, Lecture 7 discusses experimental evidence from Farfàn, McGee, Perng, and Vakis (2019) in some detail, and summarizes guidelines from Oseni, Durazo, and McGee (2017). The Instructor is encouraged to review these additional references.

Time allocation

Food away from home	
Concepts	10 min
Empirical evidence	30 min
Recommendations	10 min
Break	**15 min**
Non-standard measurement units	
Concepts and checklist for implementation	40 min
Recommendations	10 min
Lessons learned	**5 min**

Annotated lecture

The last lecture on measuring food consumption tackles two very specific topics: food away from home, and non-standard measurement units.

The first few slides define what exactly is meant by food away from home: the focus is on food that is *prepared* outside of the home, and therefore does not enter the budget of the household, unless it is explicitly accounted for. Some evidence is cited to motivate the emphasis on correctly measuring food away from home.

Evidence from an experimental study by Farfan et al. (2019) is used to showcase different approaches to the measurement of FAFH, and to establish which method performs best. Similarly to the studies conducted in Tanzania, and mentioned in previous lectures, this experiment also tests different questionnaire designs, and compares them to a 'benchmark'. The result that should be most stressed is that a single question is not enough to adequately capture FAFH. The issue, then, is how to choose the best among the remaining approaches.

A review of the current practice shows that in many cases, the measurement of FAFH is lacking. Recommendations adapted from Section 3.5 of the FAO/World Bank guidelines show how to meet the challenges related to the collection of data on FAFH.

Now is a good time for a break.

...

The second part of the lecture, on the use of non-standard measurement units, is modelled after the Guidebook by Oseni et al. (2017). The Instructor is encouraged to consult sections 3 and 4 of the Guidebook for a detailed discussion of this topic.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – FAFH modules

An in-depth analysis of the two questionnaires may raise some or all of the following points.

Both questionnaires capture the data on food consumed away from home (FAFH) with a recall period of 7 days. Tanzania has a distinct module/section dedicated to FAFH. By contrast the data is incorporated in the overall food and beverage consumption in Uganda. The major differences between the two questionnaires are twofold:

1) FAFH in Uganda is collected for the entire food list, whereas Tanzania concentrates on 7 (somewhat ambiguous) composite items (full meals, barbecued meat chips and snacks, local brews, tea coffee samosa cake and snacks, alcoholic and non-alcoholic beverages).
2) Tanzania accounts for FAFH received as gift and differentiate them from other food received for free (collected in another module). In Uganda FAFH is collected for food purchases only (as the phrasing of the question seems to indicate).

The minimum reliability criteria from Smith et (2014: 23) are met for Tanzania only. In both questionnaires the sources of FAFH received in kind are missing.

Exercise 3 – Non-standard measurement units

Picture 1 lacks a familiar reference, like a pen or a bottle, that would allow to gauge the size of the fruits. Picture 2 contains a reference, but it is not very visible, and perspective is skewed: the largest heap is closest to the camera, thus appearing even larger than its actual size. A better way to picture the units would be to frame them straight on against the backdrop. A similar problem occurs with picture 3: this time the bottle is clearly visible, but the baskets are pictured from above, and the largest appears closer to the camera, again skewing perspective and making it difficult to estimate the size of the containers.

Lecture 8
Measuring consumption of non-food non-durable goods

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on expenditures on non-food, non-durable goods through household surveys. First, the lecture gives a broad description of the analytical procedures usually adopted when analysts construct a proxy of non-food consumption; this knowledge is then used to inform data collection recommendations.

Suggested preparation

The Instructor is assumed to be familiar with the material contained in section 3 ("Constructing the Household Consumption Aggregate") of Deaton and Zaidi (2002). Guidelines on data collection of non-food non-durable expenditures are from Deaton and Grosh (2000) (general guidelines), Oseni et al. (2018) (education), Gertler, Rose, and Glewwe (2000) (health): this material should be read in advance. The lecture also cites evidence from Heijink et al. (2011) and Lu et al. (2009).

Time allocation

What analysts do, 'in or out?'	**45 min**
Break	**15 min**
What analysts need	
General guidelines	15 min
Education	20 min
Health	20 min
Lessons learned	**5 min**

Annotated lecture

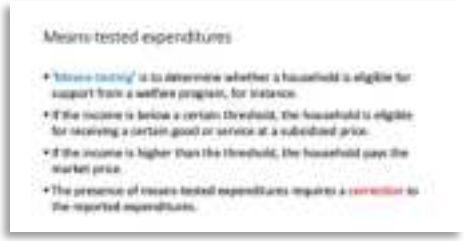
The outline clarifies the 'grassroots' approach of the lecture: the discussion starts from research objectives (what do welfare analysts do when constructing the non-food component of a measure of living standards?), and then proceeds to data collection guidelines (how can the questionnaire meet the research objectives?).

...

The takeaway from this first part of the lecture is that not all expenditures should be indiscriminately considered to be part of the living standard indicator. The *consumption* aggregate is constructed by analysts based on *expenditure* data, which requires some fine-tuning: some expenditures should not be included in the aggregate, because they do not represent *typical*, *welfare-enhancing consumption*.

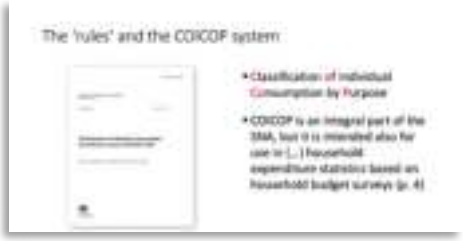
...

Next, the lecture discusses inclusion and exclusion 'rules' in further detail. Items to include are generally uncontroversial; the only categories worth a comment are regrettable necessities (see Deaton and Zaidi, p. 32 for an effective discussion of this point), and utilities (which can be problematic when, for different reasons, they are not purchased at market price: Deaton and Zaidi touch on this on p. 34).



Expenditures to be excluded are only briefly discussed: two upcoming lectures will focus specifically on durable goods and housing. The problems related to means-tested expenditures are similar to those arising in the case of subsidized items: these expenditures must be excluded, unless their price can be corrected and equalized across households.

Finally, the controversial nature of health expenditures is briefly touched on: see Deaton and Zaidi, p. 30-32 for a detailed discussion. More will be said on health expenditures in the rest of the lecture.



Students may already be aware of the COICOP classification of expenditures; if not, the availability of this system is an important takeaway of this lecture. The UN COICOP classification provides a basis that each country can extend at will, and it is often used in household consumption and expenditure survey questionnaires. The COICOP classification is used to put the 'rules' discussed in the previous slides to the test. Students should be invited to attempt to explain why certain items should or should not be included in the measure of living standards.

Food, beverages, clothing, housing and utilities should all be included. The danger sign on narcotics indicates the potential criticism of narcotics not being welfare-enhancing (but Deaton and Zaidi recommend their inclusion, nonetheless).

Expenditures in these categories are to be included, but students should be reminded of the controversial nature of health expenditures.

Vehicles are classified as durable goods. The remaining items are to be included.

If the label 'financial services' indicates savings or investment, they should be excluded. Note that insurance services are, in principle, consumption (consumers purchase a service, which is the insurance company's promise to protect them from a potential bad outcome), but in some countries, insurance is used as a form of investment. In that case, it should also be excluded.

Finally, these categories are used for national accounting, and are typically not needed when designing household surveys.

Now is a good time for a break.

Next, the lecture moves on to consider data collection recommendations, on the basis of the analytical priorities discussed so far. It should be acknowledged that guidelines on the collection of non-food, non-durable expenditure data are rather broad, and for many choices that questionnaire design teams will face, there is no single universal answer. This portion of the lecture is based on Deaton and Grosh (2000), which describes the general approach adopted by LSMS surveys.

Oseni et al. (2018) provide more specific guidelines on education. The Instructor is referred to the Guidebook for details.

A useful reference for specific guidelines on the health module is Gertler et al. (2000), which draws once again from the LSMS experience.

The lecture is wrapped up by a discussion of the current practice. The point to be made is that there is some leeway in designing many of the important features of the health module (recall period, disaggregation, structure).

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – 'In or out?'

Students must read Section 3 of Deaton and Zaidi (2000) in order to engage with this exercise. However, an exact solution is not the goal here. Not all items will be uncontroversially in or out: the exercise can be used to spur a discussion on analytical choices, and their bearing on data collection.

Exercise 3 – The treatment of health expenditures

The goal of the exercise is to let students evaluate the practical repercussions of the concepts discussed during the lecture. Both countries exclude health expenditures from the consumption aggregate; the motivation given by both is in accord with the arguments presented by Deaton and Zaidi (2000), except for the inclusion of purchased medicine in the case of Bhutan, which is essentially left unmotivated.

Exercise 4 – The education module

This exercise gives students an opportunity to see a practical application of the concepts discussed in the lecture, and become familiar with questionnaires used in different countries. The goal of the exercise is to assess the students' grasp of the guidelines on the education module, and see whether they can critique questionnaires based on the recommendations learned during the lecture.

Lecture 9
Durable goods

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on durable goods. The lecture explains how analysts treat this category of expenditures when constructing a measure of living standards; these analytical requirements guide the design of a survey module on durable goods.

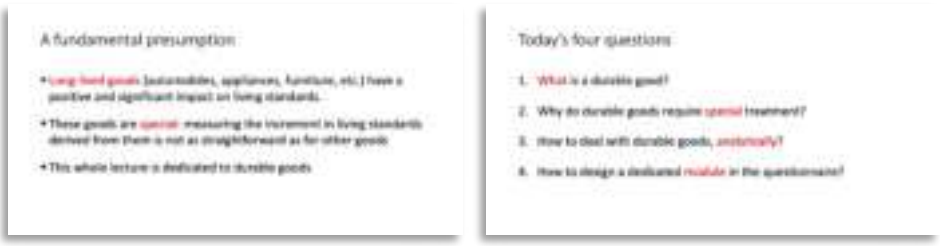
Suggested preparation

The Instructor is assumed to be familiar with Section 3.4 of Deaton and Zaidi (2000) (Consumer Durables), and with Amendola and Vecchi (2014).

Time allocation

What are durables, and why do they require special treatment?	**20 min**
How to deal with durable goods	
Notation. Acquisition and rental equivalence approaches	40 min
Break	**15 min**
User cost approach, data requirements	25 min
How to design a dedicated module in the questionnaire?	**15 min**
Lessons learned	**5 min**

Annotated lecture

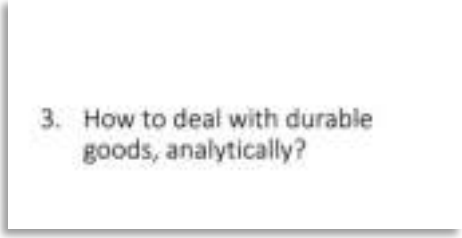


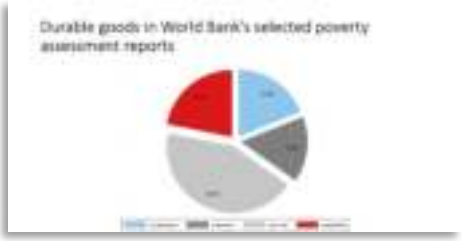
The first few slides of this set make an important point: **durable goods** are not like other consumption goods that contribute to living standards, and they **require special treatment** when we construct a consumption-based measure of welfare. *Why* this is the case and *how*

to treat these goods, both in terms of analysis and if questionnaire design, will be the subject in this lecture.

First, we give an exact **definition** of durable goods, using Diewert (2009). Typical examples of durable goods are housing (which requires separate treatment, and will be discussed in the next lecture), cars and other vehicles, household appliances, and so on.

Next, why is it that durable goods require a special treatment? The key concept is that welfare analysts want to measure consumption during the reference period, and not expenditure. For most goods, which are purchased and consumed right away, expenditure is a good proxy for consumption. But the use of a durable good exceeds, by definition, the reference period, and its purchase price reflects this long-lasting flow of consumption that the consumer is able to enjoy from the good: if you buy a bycicle today, you are paying for the consumption of it throughout its duration, which is much longer than a year. Instead, welfare analysts want to **measure only the portion of consumption that is enjoyed during the reference period**.





The next question is, how do welfare analysts treat durable goods? There are three main analytical approaches, and each of them requires different pieces of information – which in turn reflects on how the durable goods module of the questionnaire should be designed. The first approach (**acquisition**) is cited as a baseline: it consists in ignoring the peculiar nature of durable goods and the 'special treatment' due to them, and simply using the purchase value as a representation of the value of consuming them. This is obviously incorrect, given the previous discussion. The **rental equivalence** and **user cost** approaches are, instead, two ways of computing the use-value of the durable good during the reference period. The pie chart shows the relative frequency of the approaches in recent poverty assessments.

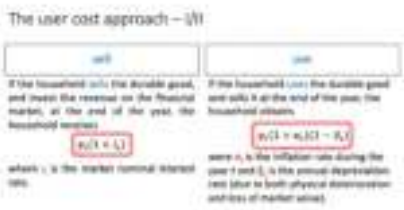
How do the different approaches work in practice? We introduce some conventinal notation and a basic formula that helps us think more clearly about the problem at hand: in a nutshell, the **consumption flow** from the durable good during the reference period, CF_t, is a fraction k of its current market price.

First, we discuss the acquisition approach – basically, *not* treating durable goods any different than any other good. If a household purchases a durable good, say a car, during

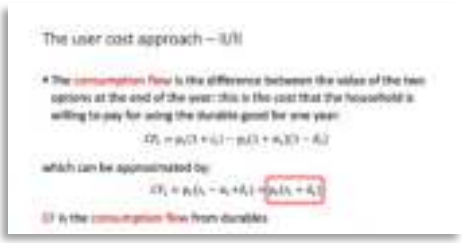
the survey period, then they are assumed to have consumed its entire value that same year; if a household did not purchase any cars during the durvey period, they are assumed to not enjoy any returns from it, *whether or not they own a car*. This is not correct from a conceptual standpoint: again, it is *use*, not *purchase* that counts for well-being.

Then, the **rental equivalence** approach. In theory, because rent as a representation of the use-value of a good for a particular time period, it is very sensible to try to measure CF_t as a rental value. The problem is, a **rental market does not exist** for most durable goods, so that this approach is generally unfeasible in practice. We will see that, unsurprisingly, this is not the case for housing. The optional (starred) slide presents the rental equivalence approach in terms of the basic equation introduced at the top of this section.

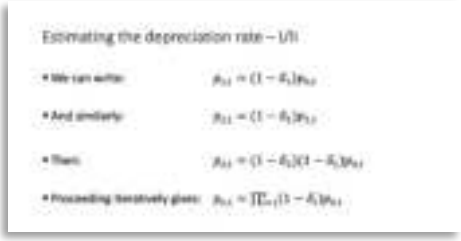
Now is a good time for a break.



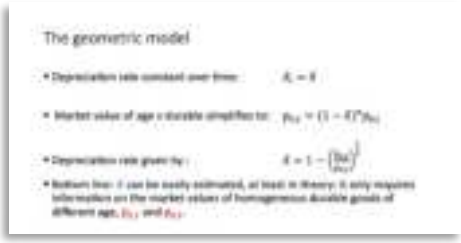
Finally, we discuss the **user cost method**, which requires a bit more time, because it is the most used in practice. The first step is to recognize that a household that owns a durable good can be seen as facing **two options** at the beginning of the survey period: to sell it right away, or to use it for a year, and then sell it.



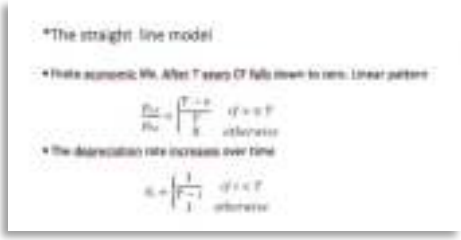
The difference between the values of the two options is an estimate of CF_t. If a household does not sell the good, but uses it, it is because the good's utility compensates for that difference. Another way to see the final equation is that, by holding the good for a year, the household is 'paying' the opportunity cost of selling it, plus the economic depreciation of the good. This can be brought back again to the general formula seen at the start.

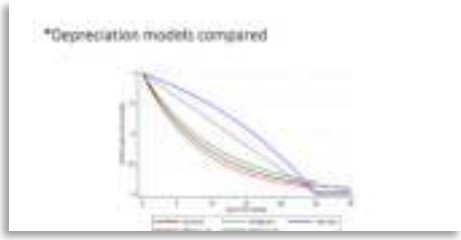


We now have a formula for CF_t, but we still do not know what the depreciation rate **delta** is. If we think about delta as relating the market value of a new good at time t ($p_{0,t}$), with the market value of a 1-year-old good of the same type, again at time t ($p_{1,t}$), then we can proceed iteratively until we reach a general expression. This expression is one step closer to knowing delta (assuming that we know both $p_{v,t}$ and $p_{0,t}$); but still, the expression says that we have a different delta for each year of the good's life, which is highly complex.



To overcome this complexity, we need to make assumptions, and to **model delta**. The most common assumption is that of a **geometric depreciation model**: this simply means that we assume delta to be the same every year. This delivers a simple way to estimate delta – at least in theory.

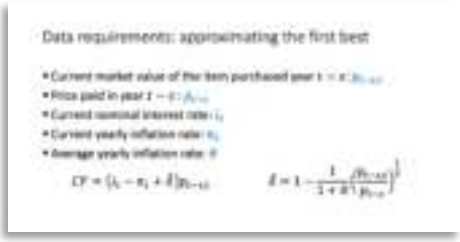




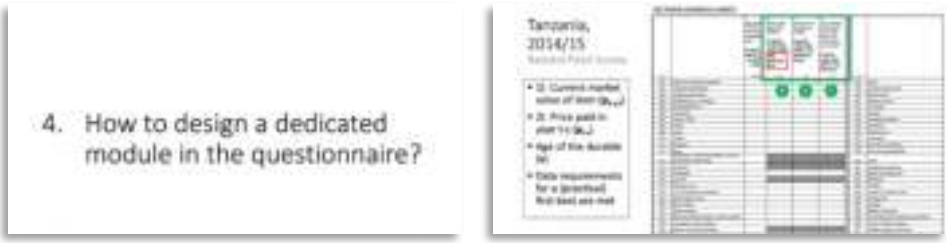
The optional slides mention alternative models that allow for an estimation of delta.

This slide gives an opportunity to recap the user cost approach, and the ‘ingredients’ that are needed to implement it: this is where data comes in. Questionnaire design must match the analysts’ needs, and gather enough information to allow for the user cost approach to be feasible in practice.

Ideally, all of the ‘ingredients’ of the formulas reproduced at the bottom of the slide (the user cost method for estimating the CF from durable goods) would need to be gathered by the questionnaire, except for the nominal interest rate and inflation rate, which are typically available from external sources. If this is the case, the analyst would simply plug survey data into the formula: this is what is meant by first best.



Realistically, obtaining enough reliable information on all of the ‘first best’ variables is not easy. In practice, a more effective (‘second best’) approach is to ask households about the purchase time, purchase value, and current market value of items they own. The years of ownership, *s*, can be used as a proxy for the ‘vintage’ or age of the durable, *v*, if needed. This still allows to apply the user cost method, provided that some changes are made to the formulas: the new expressions are reproduced below.



This is an expample of a survey that correctly gathers information needed to implement the second best approach (here called a ‘practical first best’). In this case, we know *v*, rather than *s*, which is ok from an analytical standpoint.

This, on the other hand, is an example of a survey that does not gather enough information to implement even the second best approach.

In practice, even if the ‘second best’ approach is not feasible, we can still apply alternative methods, and estimate the necessary variables in an indirect way. In a case like Palestine this is really impossible: the available data is too limited.

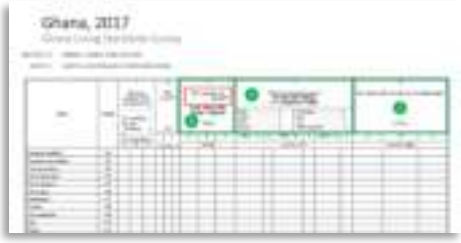
Lessons learned deserve a few minutes of our time. The takeaways are that durable goods require special treatment when constructing a measure of welfare, because welfare analysts want to proxy consumption; to do that, the user cost method is recommended. The questionnaire should definitely include a module on durable goods, and gather as much information as needed to apply the user cost method.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – The durable goods module



Data requirements for the second best are met even in the absence of the age v of the durable:

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
2) Price paid in year t-s ($\mathbf{p_{t-s}}$)
3) Years of ownership (**s**)

Note a potential problem with the question: "How long ago was […] obtained". If many durables were obtained as gifts, then the reported purchase value of goods will often be zero, and analysts may need to impute these values.

Similarly to Ghana, the data requirements for a second best are met even in the absence of the age *v* of the durable:

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
2) Price paid in year t-s ($\mathbf{p_{t-s}}$)
3) Years of ownership (**s**)

Here, information is collected on the "most recent" durable. How does this affect the resulting estimate (think about cases in which the household owns more than one durable good for each type)?

In this case, standard methods cannot be applied.

1) Current market value of item ($\mathbf{p_{t\text{-}s,t}}$)
~~2) Price paid in year t-s ($\mathbf{p_{t\text{-}s}}$)~~
3) Years of ownership (**s**)

Alternative methods exist when we only have the current market value of the durable, and an estimation of the **maximum economic life of the durable.**

The questionnaire only collects information about the price paid if the durable was acquired *in the last 12 months* (in year t!).

1) Current market value of item ($\mathbf{p_{t\text{-}s,t}}$)
2) Price paid in year t ($\mathbf{p_t}$)
3) Age of the durable (**v**)

No information is collected on the price paid in year t-s; however, note that, in this case, $\mathbf{p_t}$ and $\mathbf{p_{t\text{-}s,t}}$ will probably be very similar value. This case is similar to **Nigeria,** but the fact that there is information on the age of the durable puts the analyst in a better position to use alternative methods for estimating the CF.

Lecture 10
Housing

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on housing expenditures. The lecture explains how analysts treat this category of expenditures when constructing a measure of living standards; these analytical requirements guide the design of the survey module on housing.

Suggested preparation

The main reading for the lecture is Malpezzi (2002), which focuses on data collection. For a thorough understanding of the analytical requirements underlying the design of the questionnaire, it is also important to be familiar with Deaton and Zaidi (2002), Section 3.5, and Balcázar et al. (2017).

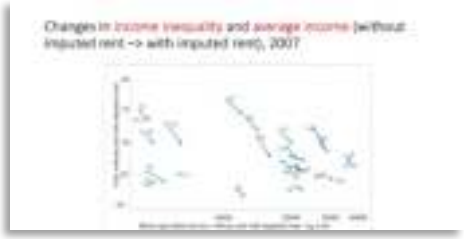
Time allocation

Concepts	30 min
Practice	30 min
Break	**15 min**
Data	40 min
Lessons learned	5 min

Annotated lecture

The set opens with a roadmap of the lecture, which is organized around three topics: concept, practice, and data.

Conceptually, housing is a durable good, and presents the same issues as those discussed in the previous lecture.



These slides present evidence of the empirical relevance of housing expenditures, in the context of welfare measurement: inequality and poverty estimates are significantly affected by housing.

Data on housing are interesting for several reasons. One is that housing characteristics are themselves an indicator of living standard: Dollar Street, a project developed by the Gapminder Foundation, is used to illustrate how dwelling characteristics go hand in hand with well-being.

Another obvious reason is that housing is a part of total household consumption, and therefore we need to account for it in the consumption aggregate.

Finally, data on housing expenditures are of interest for analysts that study the behavior of housing markets. Of course, our focus is on welfare analysis, therefore motivations 1 and 2 will drive the rest of the discussion.

Motivations 1 and 2 imply that the housing module must contain at least the pieces of information here listed. But what is the use-value of the dwelling? The next few slides explain.

Theoretical considerations here are very similar to those that apply to durable goods. Welfare analysts are interested in the **value of consumption** of durable goods, including the dwelling, not in their purchase value. What differentiates a house from other durables is the fact that **rent** is usually available, and as mentioned in the previous lecture, rent is an adequate representation of the use-value of a good. For households that rent their dwelling, the case is closed. However, households that own the dwelling do not pay any rent, of course. A similar issue arises when a household pays a rent that is not representative of the true value of their dwelling, but is artificially lowered (by subsidies, for instance). In both of these cases, what we need is the rent that owners or non-market tenants *would* pay, if they were to rent their home on the housing market. This is called **imputed rent**.

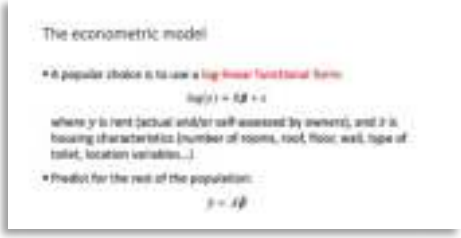
These slides reinforce the concept of imputed rent and its importance for making correct welfare comparisons. Egypt is a real-life case in which rent controls (legally enforced lowered rents for certain tenants) pose this kind of problem to analysts.

Now is a good time for a break.

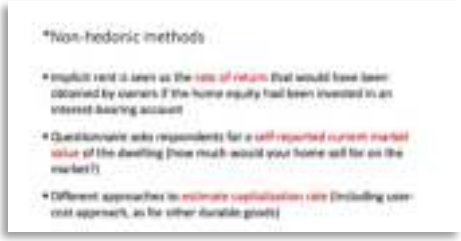
...

In practice, computing imputed rent is **not trivial**. Different methods are available. Students are directed to the article by Balcazàr et al. for a complete review. There are three main options, but we will focus on the first two.

The first approach, self-reported rent, relies entirely on the collection of data directly from respondents. This strategy is useful, and the self-reported rent question should always be included; however, there may be challenges to the quality of the data collected in this way, when respondents are not informed (the housing market in the area does not exist, or is very small, so respondents have no information to base their response on) or not objective (the perceived value of one's own dwelling is unrealistically high).



An alternative is to estimate rent based on other information, namely the characteristics of the dwelling (a nocer house would go for more on the rental market than a dilapidated one: if the analyst is able to select the characteristics that influence market rent, she can "predict" it). This approach is called **hedonic regression**, usually implemented as a log-linear specification.



These optional (starred) slides mention two extensions of the topic that was just discussed. Duan's estimator is a way to obtain more accurate predictions, and has to do with the logarithmic transformation of the dependent variable. The Instructor is invited to read the paper for more details. The second slide hints at another avaulable method besides hedonic regression, which may be used if all else fails.

Now is the time to recap the main messages of this section: the last bullet is especially important, and leads to the final part of the lecture.

These slides offer a few practical recommendations for the design of the housing module, mostly from Malpezzi (2002). First, it should be clear that concepts must be defined, however familiar they may seem.

Next, what should the housing module should ask for, in practice? Dwelling characteristics are crucial.

In terms of dwelling expenditures, the crucial pieces of information are tenure status, rent, and utilities/maintenance/repairs.

The final slides show examples of housing modules in recent HCES. Students are invited to examine them and point out which questions collect the different pieces of information mentioned during the lecture.

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Secondary residences

This exercise gives students an opportunity to explore the housing module in a number of different questionnaires, and pushes them to examine their structure in depth, by asking a question that was not covered during the lecture. Each survey will of course collect data on secondary residences in a slightly different way: rather than the answer, what is important here is the students' hands-on engagement with the questionnaire.

Exercise 3 – Housing in theory and practice

Similarly to Exercise 2, this question asks students to comment on questionnaires, this time pushing them to make a connection with the theory. Ultimately, students should be able to conclude (i) whether or not the design of the housing module allows for the implementation of at least one of the recommended methods for the estimation of the use-value of the dwelling, and (ii) what was actually done when constructing the welfare indicator.

Lecture 11
Data validation and diagnostics

Learning objectives

This lecture deals with data validation techniques and data diagnostics. The main goal of the lecture is to provide a gentle introduction to a variety of problems that arise when survey data are collected in the field and processed by the NSI prior to dissemination with final users. The lecture is closely linked to Lecture 12, which focuses on outliers, as the latter are often a specific form of (gross) errors, one of the many covered in this lecture.

Suggested preparation

The single most important reference for this lecture is the manual by de Waal et al. After becoming familiar with the sections indicated in the syllabus, Instructors are encouraged to customize the references with country specific references, as appropriate.

The number of slides in this lecture is intimidating. Many are examples, which only require a few seconds of discussion. Nevertheless, the Instructor is invited to check whether cuts are needed *before* delivering the lecture.

Time allocation

Types of errors	20 min
Data editing	15 min
Missing data	25 min
Break	**15 min**
Data validation and diagnostics	40 min
Lessons learned	5 min

Annotated lecture

We begin by ackowledging that errors are ‘part of the game’. After showing the first bullet of the first slide we can open the floor and engage with a disccusion with students, asking them what kind of errors they can think of. We do not reply, we simply ‘chair’ the discussion. Next we remind students that in this lecture we focus on errors that arise in steps 3 and 4 of the survey process.

Next we provide students with a quick overview of different types of errors. The goal is twofold: i) to introduce a few key *concepts* on the different types of error and ii) to build up students’ statistical vocabulary. This part of the lecture is a bit boring, in our experience (as are some other material down the same lecture), so we thought to provide Instructors with the option to take a *short* digression on a specific tool, the Benford’s Law, that helps motivate the audience.

...

The Benford’s Law *per se* is a topic that in this course should not take more than a couple of slides – we prepared 6. Thje reason is that it’s a nice story to tell and we use it as a device to convince students that the rest of the lecture could be as stimulating/entertaining as the Benford’s story/law. In short, the 6 slides are a rethoric device that should be used only by Instructors familiar with the references provided in the syllabus, and confident that the opportunity cost of this longer-than-strictly-needed presetnation is not too high.

...

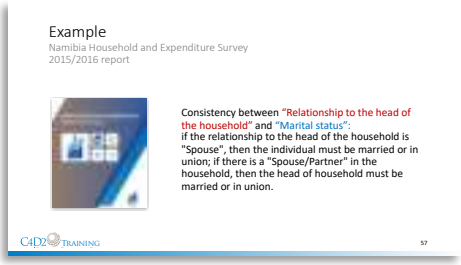
The next bunch of slides introduces and elaborates the concept of ‘data editing’. Slides are self-explanatory, we believe and can be improved by adding specific examples, if the Instructor wishes so. Our advice is emphasise and elaborate on *over-editing* – it is a common bad practice and the discussion in de Waal et al deserves some emphasis.

We kept the discussion of missing data within the spase of about a dozen slides. Given the relevance and the complexity of the topic, the choice of the contents is clearly a challenge. We decide to start with a fundamental idea: the single and most important priority when faced with missing data is to understand the *mechanism*: why are data missing in the datasets? Different answers imply different coping strategies.

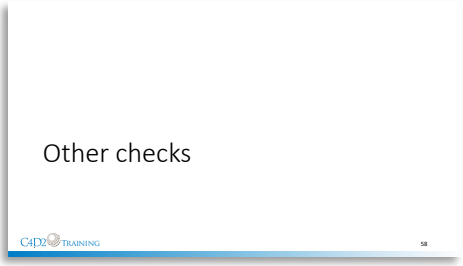
Next we illustrate and discuss the two extremes (data are MCAR, that is missing completely at random, and data are MNAR, missing not at random). We leave it to the Instructor to decide whether to engage with a formal representation of these mechanisms (a few optional and hidden slides were prepared for those willing to do so). Similarly, we leave it to the Instructor to decide whether to venture into MAR (missing at random) data. Our suggestion is to prepare this part of the lecture carefully, as time flies when you try…

The remaining slides deal with the *treatment* of missing data. To impute or not to impute? This is the question that will keep the class busy until the break, we suppose.

Now is a good time for a break.



After the break we introduce the last topic of the lecture, data validation and diagnostics. After defining what we mean by that, we provide a selection of examples that can be discussed with students interactively. "Range checks" are usually simple to cover (here a reminder of the discussion on data capture methodolgies might be in order), and so are the examples of "consistency checks". Emphasis here is not much on specific examples, but rather oin the fact the there is a need to introduce ciomplex check involving multiple variables possibily collected in different sections of the questionnaire.



In the last part of the lecture we introduce two tools that are often used as a check of overall data quality. The first tool simply consists of a visual inspection of the population pyramid, as estimated by the survey. Does it show suspicious dents or spikes, or any other anomaly? If positive … how to interpret it? If students are not too tired, this is a nice topic for discussion. The second tool focuses on the variable age and calculates, as a measure of age-heraping, the Whipple index. Despite its popularity, the interpretation of this index is controversial. The Instructor might wish to discuss pros and cons with students and end the lecture leaving it as an open issue.

The lecture closes with a plain recap of the main points covered: a) multifaceted and pervasive errors (not just probabilistic, as we as seen, but a long catalogue), b) among errors, missing data deserve special attention. As many problems as many solutions: this is how we used the second part of the lecture, discussing data validation and imputation methodologies.

We conclude with a most important message: the need to produce (and/or ask for) proper and detailed documentation of all choices made.

Homework

Exercises 1,2 and 3 - Engaging with the literature

Usual comment.

Exercise 4 – Data imputation

Instructor are encouraged to design and implement this exercise as they see it appropriate. One way is to ask students to write a short essay discussing the pros and cons of the imputation procedures described in the reports in the slide (but others can be chosen).

Lecture 12
Outlier detection and treatment

Learning objectives

This lecture is an extension of the topics discussed in Lecture 11 on data validation and diagnostics. The goal of this lecture is to provide a conceptual framework and some techniques to detect and treat outliers (*i.e.* extreme values), which are commonly found in most datasets. We will focus, in particular, on consumption expenditure.

Suggested preparation

Chapters 1 and 2 of Barnett and Lewis (1994) provide background for the topics of the lecture, and the Instructor is encouraged to review this material.

Time allocation

Definitions	**10 min**
Do outliers matter?	**25 min**
How to detect outliers	
Visual inspection	25 min
Break	**15 min**
Statistical methods	35 min
How to deal with outliers	**5 min**
Lessons learned	**5 min**

Annotated lecture

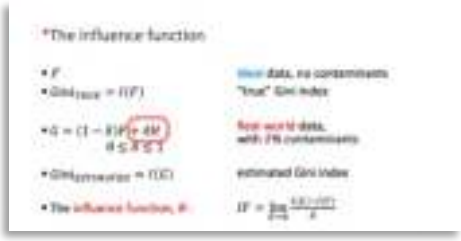
The lecture opens with a roadmap: these four questions will guide the discussion for the day.

First, the basics: what is an **outlier**? There are actually many definitions of the term, but the one by Grubbs (1969) is rather intuitive, and has become a 'classic' definition. The note in the second bullet simply means that our focus for the lecture (and the course) is on outliers within the distribution of a *single* variable; a more complex case would be that of *multivariate* outliers, which are defined based on a combination of different variables – but we will not go into that.

As will become apparent in the discussion that follows, **visualizing data** helps to perceive outliers clearly. The distributions shown in the graph (sea levels measured every year from 1930 to 1985) are concentrated around a certain range, and 'extreme' values are immediately visible. Other definitions of the concept of outlier are cited, to convey the idea that there is no hard-and-fast rule to determine what counts as an outlier, and what does not.

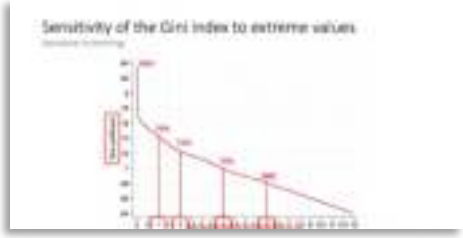
Outliers may be the result of various types of error, but they could also be **genuine** values. This is an important message to deliver: outlier is *not* synonym with 'mistake'. It is not unheard of, for variables like, say, consumption or income, to have extreme values: very rich and very poor people do exist, and they may be the very focus of our analysis. Therefore, data producers and data analysts face a dilemma: assessing whether outliers are genuine values, and, if they are not, understanding their impact on the statistics of interest.

Why should one care about outliers? The next batch of slides presents some evidence demostrating that outliers do matter in practice, in terms of their impact on outcomes of interest for data users. We will summarize the results from at least two, at most three papers (Cowell and Flachaire 2007 may be skipped, as it is quite advanced). First, Cowell and Victoria-Feser (1996): the paper is technical in nature, but its main conclusions are easy to grasp intuitively, and important to understand. The quote from page 82 of the paper is in a language that is inaccessible to most readers, but its meaning will soon be clear.

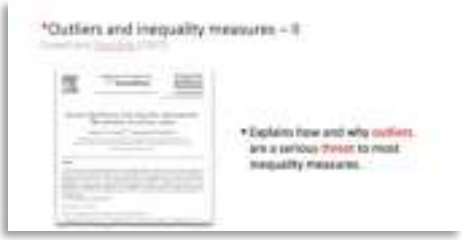


This optional slide elaborates on the concept of **influence function** (IF). The IF is a measure of the difference between the 'true' statistic of interest – in this case, the Gini coefficient – and the same statistic, computed using contaminated data (*i.e.* data containing observations that should not be part of the datasets, but that are present by mistake – they are called 'contaminants' in statistics). An unbounded IF means that no matter how few the outliers, their impact on the Gini can be infinitely large – that is, the difference between the 'true' Gini (what we would observe in the absence of outliers) and the estimated Gini (what we estimate in the presence of outliers) can be very large, infinitely large. This is why an unbounded IF is a problem.

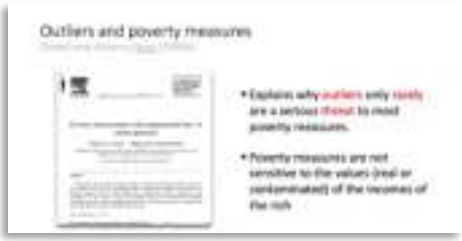
Assuming the starred slide has *not* been shown, the implications of Cowell and Victoria-Feser's findings can still be made intuitively clear. A **contamination**, that is, one or more 'wrong' observations that are far away from most other observations in the distribution, will make the distribution appear much more skewed than it actually is, according to summary statistics and inequality indicators.



A realistic example can help solidify this concept. 'Trimming' extreme values in a distribution is common practice: this just means that a researcher may drop a few of the largest observations in the sample – the top 1% households with the highest expenditure, for instance. This is similar to 'contamination': the distribution is a certain shape before trimming, and changes after. The graph shows how the estimated Gini coefficient varies when the largest observation in the distribution of per capita consumption is dropped; then the second largest; then the third largest; and so on. The horizontal axis keeps track of how many observations are dropped, as a percentage of the total sample size. As we can see, dropping just a few observations has a huge impact on Gini. There is no limit, in fact, to how large that initial drop may be, regardless of how few observations we are modifying.



These optional slides elaborate more on the impact of outliers on inequality measures. The Instructor should get familiar with the paper, in case this part is covered in class. The main takeaway is that different statistics of interest (in this case, inequality indicators) show different reactions to outliers.



To conclude this discussion, we mention another paper by Cowell and Victoria-Feser, which investigates outliers and poverty measures, and reaches a rather opposite conclusion with respect to the previous one: **poverty** measures are not sensitive to extreme values.

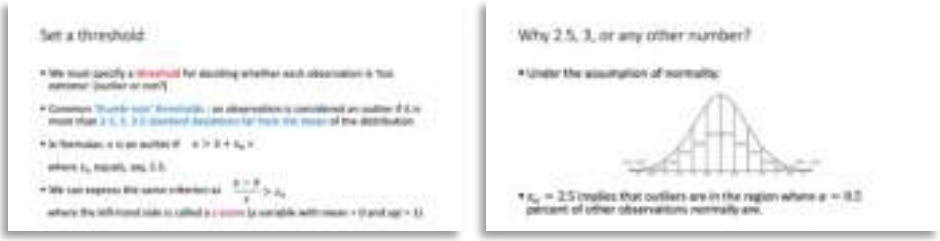
Assuming now that we do care about outliers, as they threaten at least some of the statistics of interest to the welfare analyst, how should we **detect** them? The question is not trivial, given that there is no single definition of outlier that is valid in every context. Graphical inspection or authomatic methods usually work in tandem to detect outliers.

A practical example is used to illustrate visual inspection of the distribution of expenditure. We use publicly available microdata from Malawi. Descriptive statistics, graphs of the density of the variable, and box-plot graphs are all tools that help detect any extreme values. In this case, there are two very large, clearly anomalous observations in the expenditure distribution under examination. These methods are effective in pinpointing them – but there are other methods available.

Now is a good time for a break.

These are statistical (automatic) outlier detection technique. The underlying idea is simple: we approximate the distribution of the target variable with a well-known distribution – usually, a Normal or Gaussian distribution – and then we set thresholds to identify extreme (i.e. highly unlikely) values. The easiest way to transform the variable of interest into

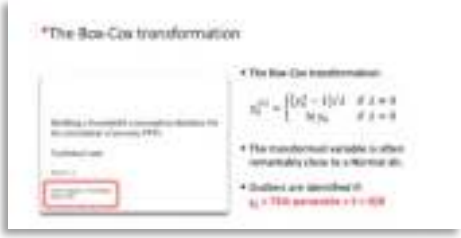
something close to a normal distribution is to take the **logarithm** of the variable, which, as illustrated, changes the shape of a skewed distribution, making it closer to a Normal (in most cases). Next, it is convenient to introduce **z-scores**. At this point, we should define the z-score, and explain that by subtracting the mean we are centering (translating/shifting) the distribution of the variable at zero, and by dividing by the standard deviation, we rescale the variable so that its variance equals one. This amounts to using a standard Normal instead of a Normal, of course.



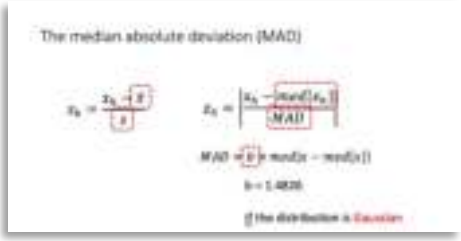
The next step is to set a threshold, after which observations are to be considered outliers. The general rule is: observation x is an outlier if $x > \bar{x} + z_\alpha \, s$, where $\bar{x}$ is the sample mean of the variable, s is its standard deviation, and z_α is a conventional value, say 2.5. The conventional values of z_α are well-known in the context of the Normal distribution: because the Normal is well known, we know that each of these values corresponds to a certain probability left in the tails.

An application of this simple procedure is in Deaton and Tarozzi (2005).

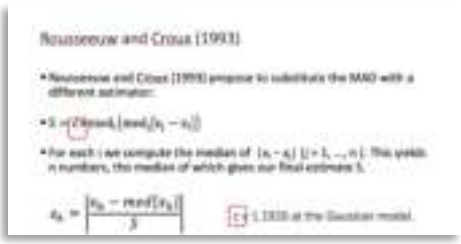
The approach can, however, be critiqued: it cannot deal with negative values (because the logarithm only exists for positive numbers), it does not apply to variables that are not normal after the transformation, and – perhaps most importantly – the thresholds that should help to identify outliers are *themselves* sensitive to outliers. As is well known, the sample mean is very much affected by extreme values, and so is the standard deviation. There are techniques that help to overcome this issue.



These optional slides hint at one of these alternatives method: it relies on another 'normaliztion' of the target variable, more complex and more flexible than the logarithm, called the Box-Cox transformation. The threshold is also different, as it relies on the interquartile range (IQR) instead of the standard deviation, and on a given percentile, for instance, the 75th, instead of the mean. These statistics are more robust to the presence of extreme values. The main downside of the method is that it is, ultimately, hard to implement, as the Box-Cox transformation cannot always be calculated. Our suggestion is to skip these two slides, unless there are specific questions on methods better than the log to induce normality in the distribution of a variable.



One solution is to 'robustify' the z-score. Instead of using a z-score based on the sample mean and the standard deviation, we replace them by plugging in robust estimates for location – using the median instead of the mean – and scale – using the median absolute deviation (MAD) instead of the standard deviation. Note that the MAD is defined as the median of all distances between each value and the median, times the parameter *b*. This is a constant that is required for MAD to be a consistent estimator of the standard deviation under the hypothesis that the distribution is Normal.



Even this approach can be improved. Following Rousseeuw and Croux (1993), we suggest to use the median for estimating the mean (numerator of the z-score) and the 'S-estimator' (denominator of the score). Uner the assumption of working with a Normal distribution, for S to be a consistent estimator of the standard deviation we need to set the parameter c equal to 1.1926. These are details that are not important for students to be confident with –

it is enough to explain why those parameter are there. Statistical softwares take care of these details.

To conclude this topic, we can say that the approach of simply taking the logarithm of the target variable and then using z-scores to detect outliers is not the best available method. Using other statistics to robustify the z-score gives better results in practice. The Malawi microdata are used to show the output of the **outdetect.ado** Stata command (Belotti and Vecchi, 2019), which performs the Rousseeuw and Croux (1993) robustification.

The final topic for the lecture is outlier treatment. There are several approaches that are routine in practice. The main message is that there is no single/best solution: the choice depends on the context.

As usual, it is good practice to take some time to end the lecture, and summarize the main takeaways. The first point should emerge strongly: documentation of each of the steps discussed during the lecture is key for any type of data cleaning or analysis.

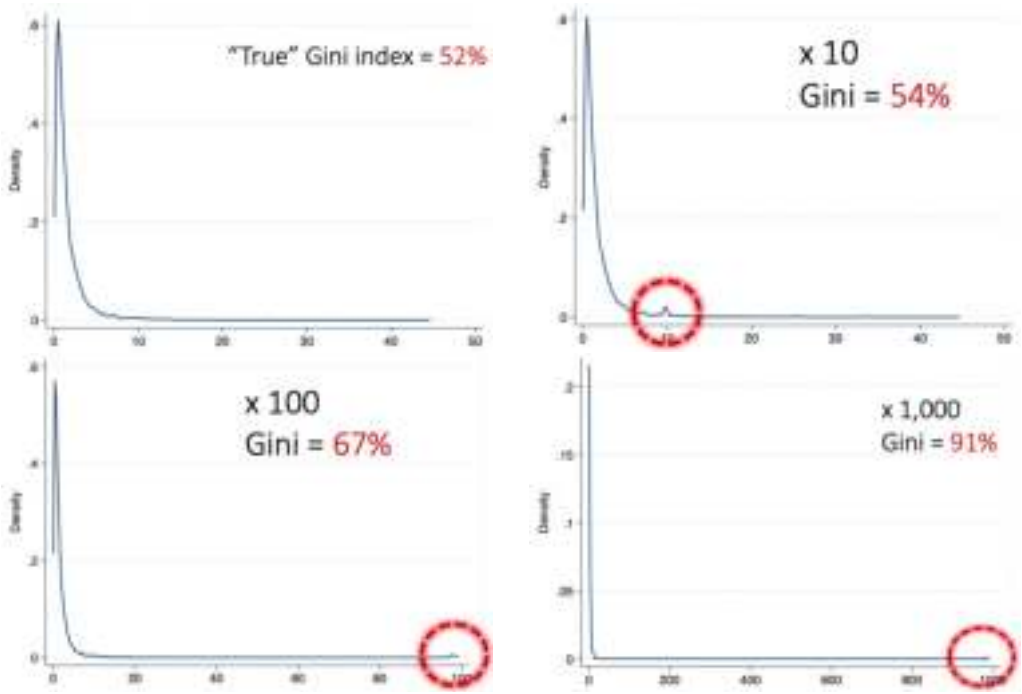
Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Do-it-yourself…

This exercise is meant to give students an opportunity to do a hands-on experiment on the impact of outliers on a statistic of interest (in this case, the Gini index), and to get a sense of how these extreme values may arise in practice (wrong placement of a decimal separator is a common data entry mistake). The code pictured in the slide is written for Stata, but the same operations can be performed with any statistical software. Results should look similar (not necessarily identical) to those pictured below:



Exercise 3 – Inequality measures

The table referenced by the exercise is an example of how strong the impact of the method chosen to treat outliers can be on final estimates. It should be stressed that none of the columns in the table report the 'correct' estimates, necessarily: they are simply the results of different ways to deal with extreme values. The importance of documenting outlier detection methods should also be emphasized.

Lecture 13
Measuring inequality

Learning objectives

The goal of this lecture is to provide an overview of both concepts and methods underlying inequality measurement. The task of estimating inequality is greatly facilitated by the availability of suitable data – thus the lectures provides the theoretical framework required for data producers and users to share a common view of what is meant by 'inequality' and how to measure it.

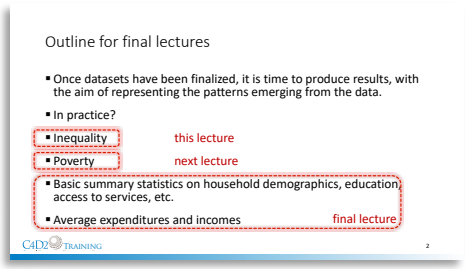
Suggested preparation

The lecture follows chapters 1 and 2 of Cowell (2011). The Instructor is assumed to be familiar with this material.

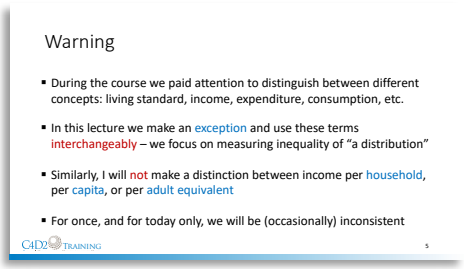
Time allocation

Tables and graphs	45 min
Break	**15 min**
Indicators	55 min
Lessons learned	5 min

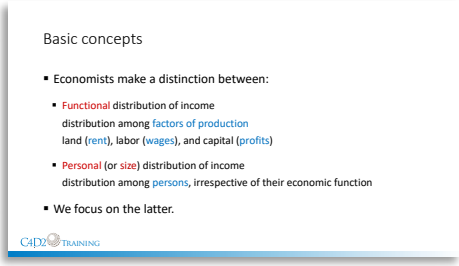
Annotated lecture

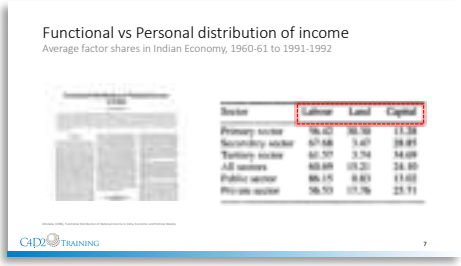


...

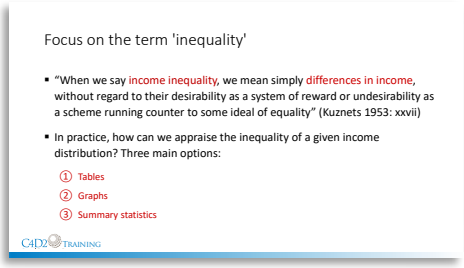


The lecture opens with a few introductory slides where we provide a short recap of where we stand, and we discuss the next steps. The main message is that today's class is the first of a twin set of *theoretical* lectures on inequality (today) and poverty (tomorrow) measurement.

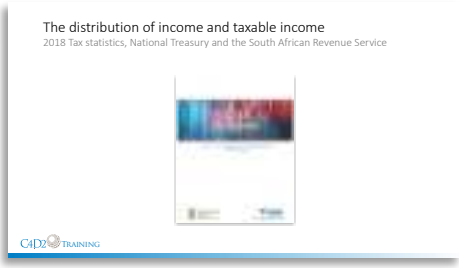


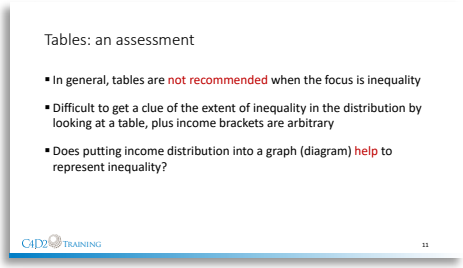


The next two slides are meant to emphasize that we will be dealing with inequality in the *personal* (or *size*) distribution of income, and dismiss the function distribution, which is typically more of interest to macroeconomists than to welfare economists.



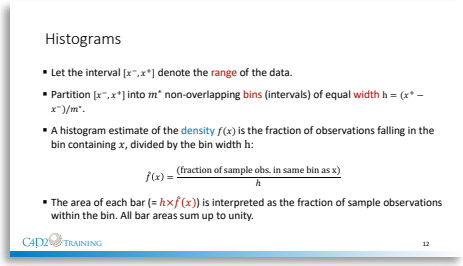
Next, we spend a few words on what we mean by 'inequality'. The point here is to clarify that in this lecture we aim at defining and measuring inequality from a purely statistical standpoint, that is, we do *not* discuss a number of interesting issues/questions related to fairness and equity – just 'differences in income' as in the Kuznets' quote. The bottom part of the slide is the outline for the rest of the lecture. The plan is to answer the question: what is the best way for describing inequality? Using tables? Graphs? Indicators?

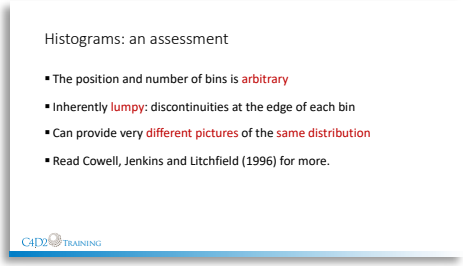




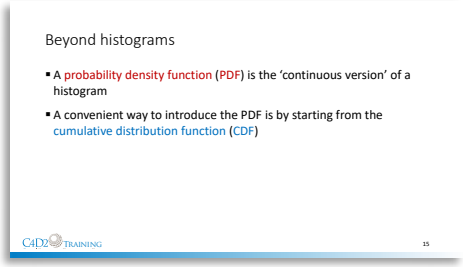
We show the first two slides to the audience and ask the question: is this table effective in conveying the extent of poverty? We engage with students and listen to their opinions. After weighing pros and cons, we argue that tables are *not* the best way to summarize inequality.

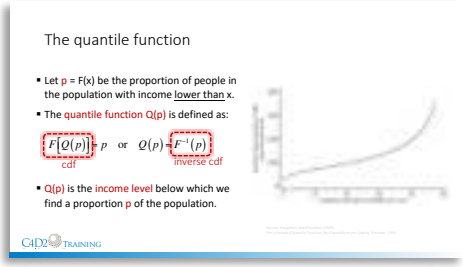
If tables are not good enough to summarizing inequality, what about graphs? Again, we briefly engage with students and ask their opinion? Which graphs would they use? Why? We listen, we do not reply and/or comment on each question. Then we take students through the next slides where we assess a selection of popular graphs that have been proposed and that people use in their work. We start with the histogram.



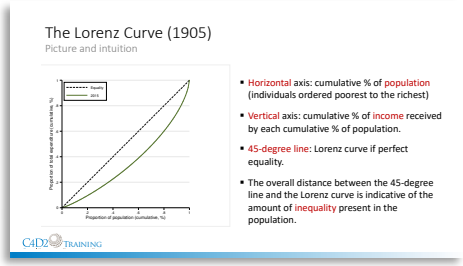


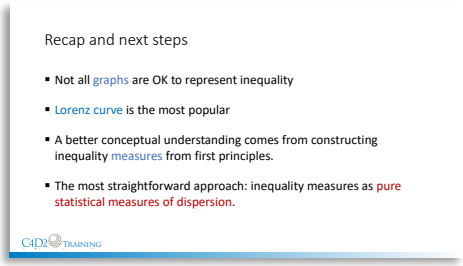
The aim of first slide is to check with the students whether they understand histograms as we think they should. Would they be able to explain how to construct a histogram from scratch? Faced with a histogram, would they be able to answer any technical question (e.g. how to interpret the height of a bar, how many bins to choose, etc.)? The message: take this opportunity to refresh your knowledge of these basic graphic tools. Only by understanding the details they can answer the question on which is the best histogram out of the three shown in the slides. After discussing pros and cons of histograms we conclude that, despite their popularity, they do not provide a 'good' way of summarizing inequality. The next three candidates are the probability density function (PDF), the cumulative distribution function (CDF), and the quantile function.





Following a similar approach, we start by introducing the CDF and conclude that it is probably a good way for getting clues about *poverty*, but not for inequality. We introduce the PDF and conclude that it ultimately suffers from the same limitations as the histogram. Next, we introduce the quantile function, the inverse function of the CDF. We concede that perhaps it is not totally ineffective at summarizing inequality, and if time permits, we explain why by means of the parade of dwarfs. We conclude that we hope to find better graphical tools. This is time for introducing *the* graph that – so we argue – succeeds in representing inequality: the Lorenz curve.



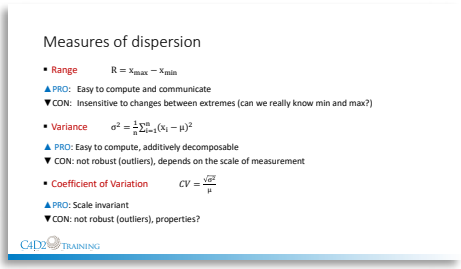


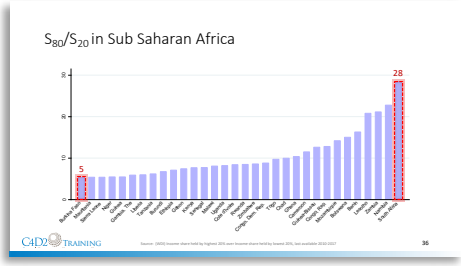
We discuss the Lorenz curve at some length and conclude that this is a useful tool, one that helps welfare analysts tremendously. This must become part of every student's analytical toolbox.

With the Lorenz Curve we conclude the review of graphical tools for analysing inequality and this is when we are likely to need a break.

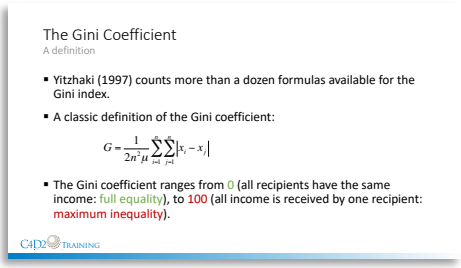
Now is a good time for a break.

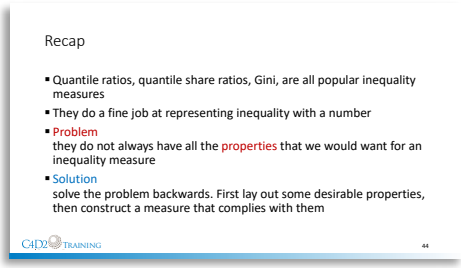
The second part of the lecture is devoted to inequality indicators.



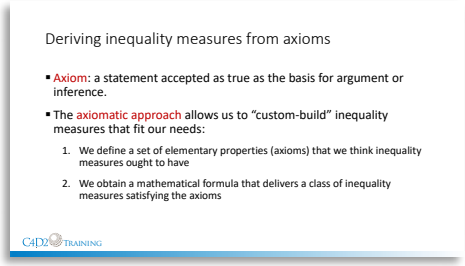


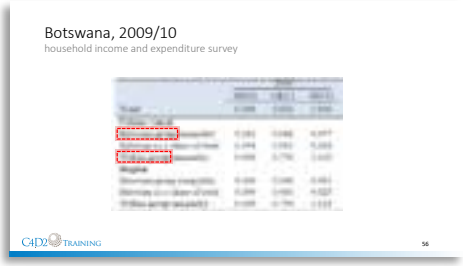
The first bunch of slides is introduced by asking students: how would you summarize, with a single, scalar indicator inequality? Would you consider using the range? And we listen to their replies. What about the variance? And we listen to their replies. The bottom line here is to convey the message that there is a multitude of indices that are available, all of them with some advantages … but a larger number of disadvantages. After introducing the indices selected in the slides, we emphasize the importance of the Gini index and explain it in detail.



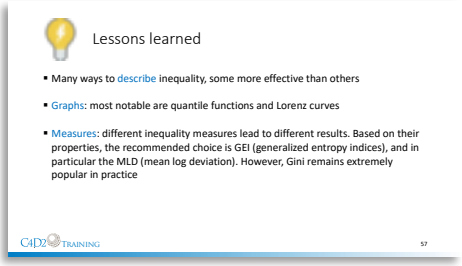


After discussing Gini, we invite students at getting as familiar with the Gini coefficient as possible, due to its wide practical application. We emphasize that, after the Lorenz curve, this is the second analytical tool that everybody should have in her toolbox.





The last bunch of slides touches on inequality indices that are derived by an axiomatic approach. This approach has revolutionized the way inequality is measured as it led to the Generalized Entropy class of Indices (GEI) that, together with the Gini index, are the most commonly used by scholars the world over. The instructor is advised to read Cowell's highly accessible account of the axiomatic approach, which underlies the slides. The take-away for students is: GEIs are important, and in particular the Mean Log Deviation (MLD).



Lessons learned can be customized, as required. Here we summarize a few key messages, namely that there are many ways to describe inequality, but only few of them 'work', that is only few have sound theoretical foundations and a wide practical application. In essence, our recommendation is to focus on the Lorenz curve (and possibly the quantile function), the Gini coefficient, and the Mean Log Deviation.

Homework

Exercise 1 - Engaging with the literature

Usual comment.

Exercise 2 – Inequality in South Asia

The main point here is that it is not a good idea to mix expenditure- and income-based inequality estimates.

Exercise 3 – Functional vs. Personal Income distribution

The exercise is for students interest in this specific topic (functionl vs. personal distribution), and as a way to check whether their understanding of the report has benefitted from attending the lecture and read Cowell's reference.

Lecture 14
Measuring poverty

Learning objectives

The goal of this lecture is to provide a conceptual framework for measuring poverty. Given the dimension and the technical contents of the literature, we limit the discussion to a short review of how analysts define a poverty line, and how they measure the incidence, depth and severity of poverty.

Suggested preparation

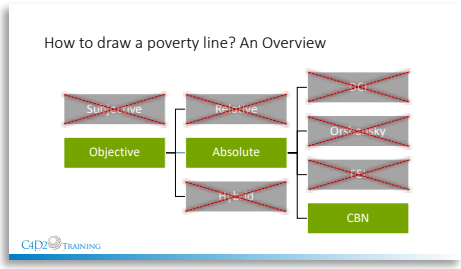
The lecture follows the sections indicated in the syllabus from Ravallion (2016). The Instructor might find it useful to become familiar with the material in Chapters 2-4 in Haughton and Khandker (2009).

Time allocation

Poverty lines	55 min
Break	**15 min**
Poverty measures	45 min
Lessons learned	5 min

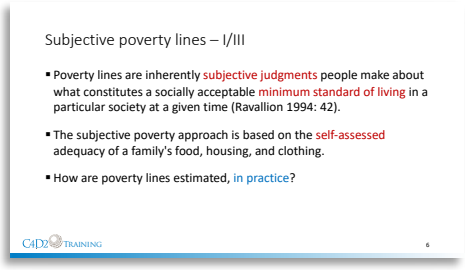
Annotated lecture

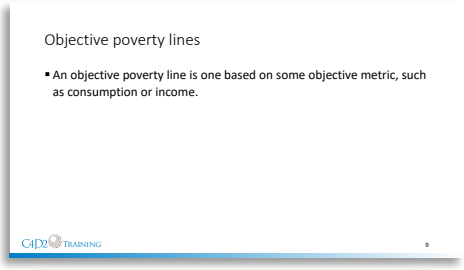
The lecture begins, as usual, with the outline of the topics that will be covered. Today is a simple two-part structure, where we explain poverty lines in part 1, and poverty measures in part 2.



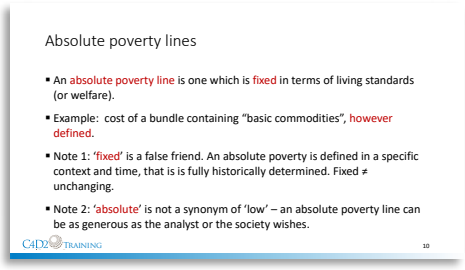
The opening slide aims at arguing that a) there is a variety of approaches that have been proposed to 'draw the line', and b) we have chosen one, out of the many possible, which is called the Cost-of-basic-needs (CBN) method.

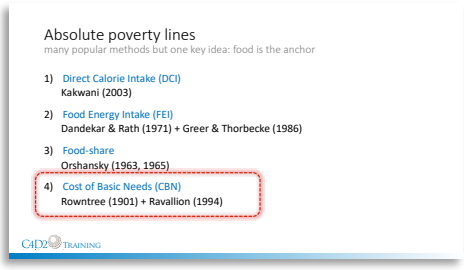
The illustration of the CBN method will take place after dismissing all other methods. We share with the audience our strategy. We announce from the onset, for instance, that we will define and explain what is meant by a 'subjective' poverty line and an 'objective' one, and explain that we will criticize and dismiss the former and focus on the latter. Next, we define and explain what is meant by 'relative', 'absolute' and 'hybrid' poverty lines, and explain that we after discussing pros and cons associated to each option we will focus on absolute poverty lines. And so on, until we reach the 'CBN'.



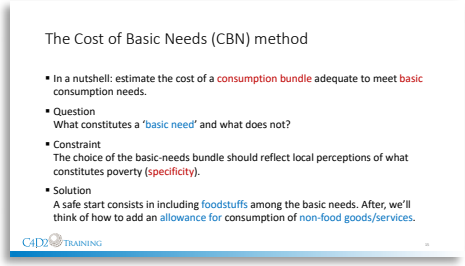


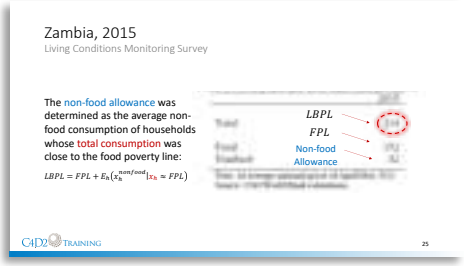
The first set of slides is on the subjective vs. objective poverty debate. The discussion is only supposed the share the *ideas* underlying these methods, to conclude that there is revival in the interest of this methodology, but currently is not the one implemented in national studies of poverty in most countries in the world. We leave it to interested students to read more on the topic.





The discussion on absolute vs. relative poverty is an important one. The key message in these slides is that relative poverty is – de facto – a measure of inequality. This message can be softened and nuanced, but this is currently the first bring-home conclusion from this set of slides. Secondly, is the idea 'absolute poverty' is not a synonym of 'extreme deprivation', nor is it an expression that does not keep the socio-economic context in due account.



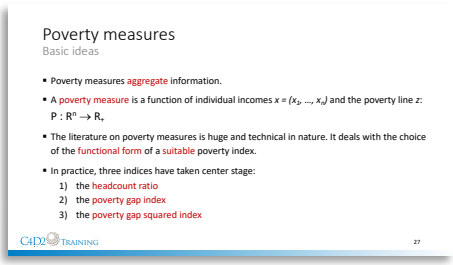


Our account of the CBN method is masterly explained in many of Martin Ravallion's work, from his 1994 book (check the references) all the way to his recent 2016 book. There is no

need to repeat here. The instructor is encouraged to customize these slides the way she/he thinks best for the audience. In our experience, the most effective way is first to explain the idea in words, next to use the support of one or more graphs, and finally to double check the understanding of the audience by illustrating the formulae (we provide the non-parametric version of the method, but the Instructor might wish to replace it with its parametric, regression-based counterpart. This is what students find in the Appendix of Ravallion (1994).

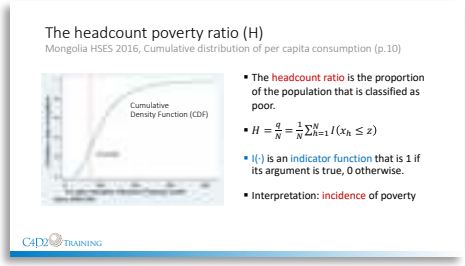
Now is a good time for a break.

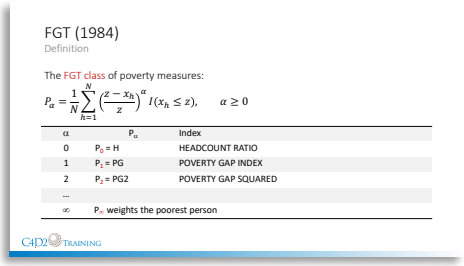
After the break the lecture resumes by the question: after estimating a poverty line, how do we summarize the extent of poverty in the society? We engage a short discussion with students. For instance, we can claim that 'the best way for measuring poverty is to count the number of households (or individuals?) whose expenditure (or income?) falls short of the poverty line. The we ask: can anyone think of a 'better' way to measure poverty? What's wrong with counting the poor? We keep the discussion short by not replying to suggestions – the point of this discussion is to deliver a first message: there are many measures available, different indicators deliver different results, not clear how to identity the best indicator. Hopefully, students should be motivated to attend the second part of the lecture.



With the first slide we repeat, by and large, the message above, but we add a good news, namely that the complexity of the choice can be reduced, in practice, to the discussion of only three indices, which we denote here by H, PG and PG2.

The rest of the lecture is devoted to introducing and assessing the three indices.

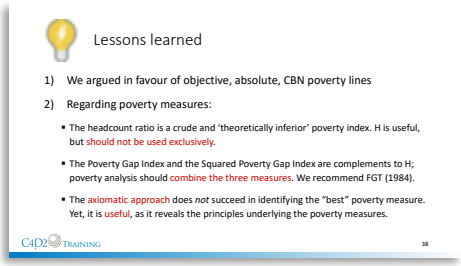




For each index, the strategy proposed in the slide is as follows. First, we provide a definition in words ("The headcount poverty rate is defined as…"). Next, we give the same definition using a formula. The notation is consistent for the three indices and the use of the indicator function (defined in the first slide) is introduced and described as a useful device to filter out the 'rich', that is, individuals whose expenditure is greater than the poverty line. We help students to read the math, exactly as if they would read plain English – this is a challenge for the Instructor, sometimes, but a high-return investment. In our experience students will appreciate mastering the formulas that we have selected. Third we provide a visual illustration of each index (with the exception of the poverty gap squared index, for which none is available). So, the headcount ratio is illustrated by means of the CDF, while the PG by means of the quantile function. Finally, we provide a short assessment of each index, a phase that lends to engaging with students (time permitting).

The material covered in these slides is pretty standard, and the Instructor should feel free to amend it by adding numerical examples, for instance.

The last two slides introduce the FGT class of poverty measures. It is important that students familiarize with this formula. The paper is a famous and classical one, but it is *not* a recommended reading for our students. Too technical and not worthy of the effort for the intended audience of this course. It is important however that students understand that i) H, PG and PG2 belong to the same family of indices (the so called 'FGT class of poverty indices'), ii) different indices assign different weights to different people (this is what we explain in the slide for PG2), 3) the FGT class of poverty indices has become the most popular among welfare analysts thanks to its theoretical axiom-derived properties.



The last slide summarizes the main points discussed during the lecture. We have been very parsimonious in summarizing the discussion on poverty lines, for instance. The Instructor can consider adding a second slide with additional conclusions/considerations, should she wish so.

Homework

Exercise 1 - Engaging with the literature

Usual comment.

Exercises 2 and 3 – DASP and ADePT

Students who have basic command of Stata can take advantage of explroing DASP and ADePT, two popular tools in use among welfare analysts. The Instructor can devise and assign specific exercises after preparing suitable small-size datasets.

Lecture 15
Describing data

Learning objectives

This final lecture closes the course by discussing the final stage of the survey process: the dissemination of final results. It presents general principles that guide the presentation of findings from a typical household survey, with emphasis on how to design effective tables and graphs, and how to present inequality and poverty estimates.

Suggested preparation

Glewwe and Levin (2005) is a useful reference on the presentation of descriptive statistics from household survey data, and the Instructor is assumed to be familiar with it. Schwabish (2014) is focused on the visualization of economic data, and is the source for some of the material presented in the lecture.

Time allocation

Background information	20 min
Descriptive statistics	40 min
Break	**15 min**
Poverty and inequality measures	45 min
Lessons learned	5 min

Annotated lecture

The subject of this lecture is the construction of the final report that is typically released in the final stages of the survey process, to disseminate the main findings from the survey, and potentially accompany the dissemination of microdata files for public use. An example of this kind of report is reproduced here (cover of the Rwanda 'Main indicators report') .

To put things into context, this slide lists some of the benefits, risks and costs related to the dissemination of findings, and, potentially, of datasets. There is no need to spend too much time on this slide: it is meant to convey the idea that disseminating results is an important task, but that it must be managed properly.

The rest of the lecture will present some guidelines that can be applied to the construction of the final report. Context matters, of course, but it is possible to make some general points. We focus on three main components of the report, listed here.

...

Some background information on the survey process should be included in any report. These slides present a 'checklist' for the pieces of information that should never be missed.

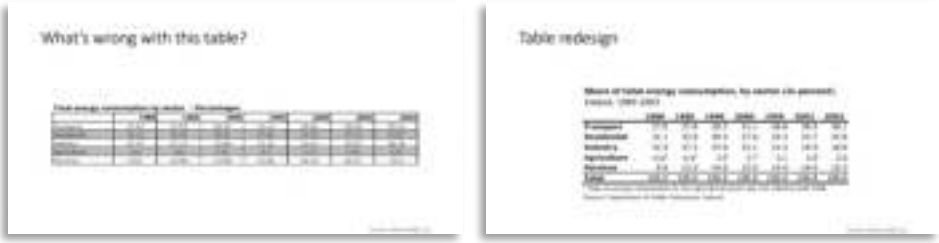
...

We now go into further detail, by providing some concrete examples. These slides illustrate the type of information on sampling design, data collection, and dataset construction that may appear in a typical report: there is no need to dwell on each detail, usually experts will be tasked with drafting these technical sections. However, it is important for students to

understand what is meant by ‘documentation’, in practice. Examples from recent reports – Kenya, Uganda, and Egypt – have the same illustrative purpose.

Next, we delve into the topic of presenting descriptive statistics. Text, tables, and graphs are our main tools when it comes to describing data; text, that is, writing about data, is a topic in and of itself and is not going to be discussed here. Interested students are encouraged to check the cited reference.

We offer a few ‘golden rules’ on how to present effective tables: these boil down to two simple but crucial principles: explaining contents clearly, and reducing clutter. We elaborate on each ‘rule’ by offering specific tips.



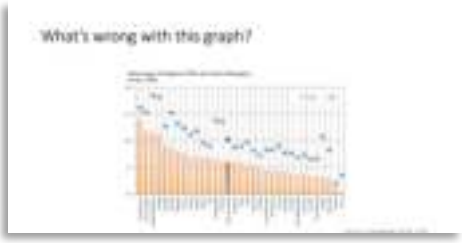
An example clarifies even further. Students can be engaged directly in commenting the first table. The ‘redesigned’ table incorporates a number of improvements: a more informative title, alignment and formatting that facilitate legibility, only one decimal digit, column totals to clarify the interpretation of cells, notes and sources.

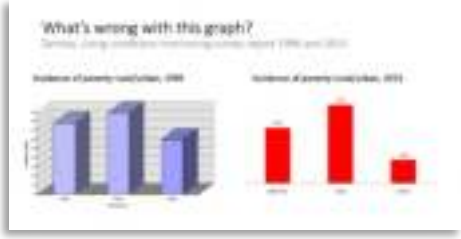
We adopt a similar approach for the case of graphs. The two ‘golden rules’ are the same in spirit, but they are obviously adapted to the context of data visualization.

Here are two common examples of graphics that do not express their message in the clearest way. Bar charts with nonzero baseline and 3D pie charts are misleading, and should be avoided.

The golden rule of ‘reducing clutter’ applies to graphs as well.

Pie charts, even when not 3D, can be a less-than-optimal way of presenting information, as mentioned by Schwabish (2014).





These slides show some examples of cluttered visualizations, and of some ways in which they can be modified to bring the data back into the spotlight, eliminating distractions.

Now is a good time for a break.

Finally, some guidelines on the presentation of results on inequality and poverty, one of the main outcomes of consumption and expenditure surveys.

• • •

Poverty and inequality measurement have been the subject of previous lectures. At this stage – the dissemination of results – it is useful to know which of the available measures have come to be expected by the international community of readers, practitioners, and academics, who are likely in the target audience of the report. The **Gini index and FGT poverty measures** are essential; more measures are a welcome addition. This group of slides includes examples from recent reports.

Another important point is that of **comparability**. Methodological changes are a threat to consistent comparisons over time and across countries, and it is crucial that they are thoroughly documented when results are released, to allow for a critical assessment of trends.

A final point concerns **uncertainty**. One of the recurring topics of the course has been the impact of survey design features – as well as analytical choices – on final estimates. These slides emphasize the importance of being transparent in conveying the unavoidable uncertainty of estimates, and recommends two tools, which are too rarely used in standard reports: standard errors, and sensitivity analysis. Examples from recent publications illustrate.

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 - Standard Errors

This exercise asks students to elaborate on the meaning and importance of reporting standard errors. One way to discuss the tables is the following. The point estimate plus or minus 1.96 times its standard error gives the upper and lower bounds a 95% confidence interval. For instance, the confidence interval for the poverty headcount rate in urban areas is (27.5, 39.3), obtained as 33.4 +/- 1.96*3. If one were to repeat this calculation for all reported point estimates, one would discover that at least some of the confidence intervals

for strata overlap, which would imply that, for instance, poverty rankings of geographical areas may need to be taken with caution.

Exercise 3 - Sensitivity analysis

This exercise is aimed at conveying the meaning of sensitivity in a practical setting. Changing the minimum calorie requirement from 2100 to 2200 kcal per person per day causes a large change in the poverty line (green to blue line), and, consequently, in estimated poverty. Adding just 23 more calories to the calorie norm generates a jump in the line that is almost as large (blue to grey line), while increasing the norm to 2300 only moves the line minimally. The poverty line is much more sensitive to the calorie norm at around 2200, while it becomes more robust over 2238.

References

A'Hearn, B., Amendola, N., and Vecchi, G. (2016). On historical household budgets. Rivista di Storia Economica, 32(2), 137-176.

Alkire, S. (2013). Towards frequent and accurate poverty data.

Alkire, S., and Samman, E. (2014). Mobilising the household data required to progress toward the SDGs.

Alvarez, E., Garcıa-Fernández, R. M., Blanco-Encomienda, F. J., and Munoz, J. F. (2014). The effect of outliers on the economic and social survey on income and living conditions. World Acad. Sci., Eng. Technol., Int. J. Soc., Behav., Educ., Econ., Bus. Ind. Eng, 8, 3276-3280.

Amendola, N. and G. Vecchi (2014), Durable goods and poverty measurement, World Bank Policy Research Working Paper no. 7105.

Atkinson, A. B. (1970). On the measurement of inequality. Journal of economic theory, 2(3), 244-263.

Atkinson, A. B. (2015). Inequality: What Can Be Done? Harvard University Press.

Atkinson A. B., WBG (2017). Monitoring Global Poverty, Report of the Commission on Global Poverty.

Backiny-Yetna, P., Steele, D., & Dijma, I. (2017). The impact of household food consumption data collection methods on poverty and inequality measures in Niger. Food Policy, 72, 7-19.

Balcázar, C. F., Ceriani, L. , Olivieri, S. and Ranzani, M. (2017), Rent-Imputation for Welfare Measurement: A Review of Methodologies and Empirical Findings. Review of Income and Wealth, 63: 881-898.

Barnett, V., and Lewis T. (1994). Outliers in Statistical Data. 3rd edition. J. Wiley and Sons.

Barton, A. J. (1958). Asking the embarrassing question. Public Opinion Quarterly, 22, 67-68.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. Journal of Development Economics, 98, 3–18.

Bouis, H., Haddad, L., and Kennedy, E. (1992). Does it matter how we survey demand for food?: Evidence from Kenya and the Philippines. Food Policy, 17(5), 349-360.

Borlizzi, A., Delgrossi, M. E., and Cafiero, C. (2017). National food security assessment through the analysis of food consumption data from Household Consumption and Expenditure Surveys: The case of Brazil's Pesquisa de Orçamento Familiares 2008/09. Food policy, 72, 20-26.

Brzozowski, M., Crossley, T. F., and Winter, J. K. (2017). A comparison of recall and diary food expenditure data. Food Policy, 72, 53-61.

Burgess, T. F. (2001). A general introduction to the design of questionnaires for survey research. Leeds: University of Leeds.

Chen, S., and Ravallion, M. (1996). Data in transition: Assessing rural living standards in southern China. China economic review, 7(1), 23-56.

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Cowell, F., and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. Journal of Econometrics, 141(2), 1044-1072.

Cowell, F.A., Jenkins, S.P., Litchfield, J. (1996): The Changing Shape of the U.K. Income Distribution: Kernel Density Estimates. In: Hills, J. (Ed.): New Inequalities. The Changing Distribution of Income and Wealth in the United Kingdom. Cambridge University Press, Cambridge.

Cowell, F., and Victoria-Feser, M. (1996). Robustness Properties of Inequality Measures. Econometrica, 64(1), 77-101.

Cowell, F., and Victoria-Feser, M. (1996). Poverty measurement with contaminated data: A robust approach. European Economic Review, 40(9), 1761-1771.

Dang, H. A., Jolliffe, D., and Carletto, C. (2018). Data Gaps, Data Incomparability, and Data Imputation. Ecineq WP, 456.

De Waal, T., Pannekoek, J., and Scholtus, S. (2011). Handbook of Statistical Data Editing and Imputation. New York: John Wiley and Sons.

De Weerdt, J., Beegle, K., Friedman, J., and Gibson, J. (2016). The challenge of measuring hunger through survey. Economic Development and Cultural Change, 64(4), 727-758.

Deaton, A. (1997). The Analysis of Household Surveys: A Microeconomic Approach to Development Policy. Washington, D.C.: World Bank.

Deaton, A. , and Grosh., M. (2000), "Consumption." In M. Grosh, and P. Glewwe eds., Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience. Washington, DC: World Bank.

Deaton, A., and Muellbauer, J. (1980). Economics and consumer behavior. Cambridge University Press.

Deaton, A., and Tarozzi, A. (2000). Prices and poverty in India. Princeton, July.

Deaton, A. and Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. LSMS Working Paper No. 135. Washington, DC: The World Bank.

Diewert, W. E. (2004), "Durables and User Costs" in ILO, Consumer Price Index Manual: Theory and Practice, chapter 23, ILO / IMF / OECD / UNECE / Eurostat / World Bank.

Diewert, W. E. (2009), "Durables and Owner-Occupied Housing in a Consumer Price Index" in W. E. Diewert , J.S. Greenlees and C.R. Hulten (eds.), Price Index Concepts and Measurements, University of Chicago Press.

Dupriez, O. (2007). Building a household consumption database for the calculation of poverty PPPs. Technical note. Available at: http://go. worldbank. org/4YG7I5RGT0.

FAO and The World Bank. 2018. Food data collection in Household Consumption and Expenditure Surveys. Guidelines for low- and middle-income countries. Rome.

Farfan, G., McGee, K. R., Perng, J., and Vakis, R. (2019). Poverty Measurement in the Era of Food Away from Home: Testing Alternative Approaches in Vietnam. Policy Research Working Paper Series 8692, The World Bank.

Farris, F. A. (2010). The Gini index and measures of inequality. The American Mathematical Monthly, 117(10), 851-864.

Fellegi, I. P., and Holt, D. (1976). A systematic approach to automatic edit and imputation. Journal of the American Statistical association, 71(353), 17-35.

Fiedler, J. L. and Mwangi, D. M. (2016). Improving household consumption and expenditure surveys' food consumption metrics: developing a strategic approach to the unfinished agenda. IFPRI

Friedman, J., Beegle, K., De Weerdt, J. and Gibson, J. (2017). Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania. Food Policy, 72: 94–11.

Finn, A. and Ranchhod, V. (2017). Genuine Fakes : The Prevalence and Implications of Data Fabrication in a Large South African Survey. Published by Oxford University Press on behalf of the World Bank.

Foster, J., J. Greer, and E. Thorbecke (1984). A Class of Decomposable Poverty Measures. Econometrica, 52, 3: 761–65.

Gaddis, I. (2016). Prices for poverty analysis in Africa. The World Bank.

Gertler, Paul J., Elaina Rose, and Paul Glewwe. (2000), "Health." In M. Grosh, and P. Glewwe eds., Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Washington, D.C.: World Bank.

Gibson, J. (2007). A guide to using prices in poverty analysis. World Bank, Washington, DC.

Gibson, J., Beegle, K., De Weerdt, J., and Friedman, J. (2013). What does variation in survey design reveal about the nature of measurement errors in household consumption?. The World Bank.

Gibson, J., and Rozelle, S. (2002). How elastic is calorie demand? Parametric, nonparametric, and semiparametric results for urban Papua New Guinea. Journal of Development Studies, 38(6), 23-46.

Gibson, R. S. (2005). Principles of nutritional assessment. Oxford university press, USA.

Glewwe, P. (2005). Chapter III: Overview of questionnaire design for household surveys in developing countries. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), Household surveys in developing and transition countries. New York, NY: United Nations.

Glewwe, P., and Levin, M. (2005). Presenting simple descriptive statistics from household survey data. In UN, Household Sample Surveys in Developing and Transition Countries. Studies in Methods Series F No. 96.

Grosh, M., and Glewwe, P. (1998). Data Watch: The World Bank's Living Standards Measurement Study Household Surveys. The Journal of Economic Perspectives, 12(1), 187-196.

Grosh, M. and Glewwe, P. (2000). Designing Household Questionnaires for Developing Countries, Lessons from 15 years of Living Standards Measurement Study, Volume One: World Bank.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics, 11(1), 1-21.

Harrison, D. E., and Krauss, S. I. (2002). Interviewer cheating: Implications for research on entrepreneurship in Africa. Journal of Developmental Entrepreneurship, 7(3), 319.

Haughton, J. and Khandker, S. R. (2009). Handbook on poverty and inequality. Washington, DC: World Bank.

Heijink, R., Xu, K., Saksena, P., and Evans, D. (2011), Validity and comparability of out-of-pocket health expenditure from household surveys: a review of the literature and current survey instruments. Geneva: World Health Organization, 28.

Hentschel and Lanjouw (2000), "Household welfare measurement and the pricing of basic services", Journal of International Development, 12: 13-27.

Heston, A. and A.O. Nakamura (2009), Questions about the equivalence of market rents and user costs for owner occupied housing, Journal of Housing Economics, 18, 273—279.

Hlasny, V., and Verme, P. (2018). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data. Econometrics, 6(2), 30.

Iarossi, G. (2006). The power of survey design - a user's guide for managing surveys, interpreting results, and influencing respondents. Washington, DC: World Bank.

Ibarra, G. L., Mendiratta, V., and Vishwanath, T. (2017). Rental regulation and its consequences on measures of well-being in the Arab Republic of Egypt. The World Bank.

Jolliffe, D. (2001). Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. Journal of Economic and Social Measurement, 27(1, 2), 1-23.

Judge, G., and Schechter, L. (2009). Detecting problems in survey data using Benford's Law. Journal of Human Resources, 44(1), 1-24.

Krosnick, J. A., Et Presser, S. (2010). Question and questionnaire design. In J. D. Wright Et P. V. Marsden (Eds.), Handbook of survey research (second edition) (pp. 263-313). Bingley, UK: Emerald Group.

Lanjouw, P. (2012), "Consumption-Based Measures in Developing Nations. Lessons from Brazil", in Besharov and Couch (eds.), Counting the Poor. New York: Oxford University Press. Ch. 13.

Lipton, Michael and Ravallion, Martin (1995). "Poverty and policy," Handbook of Development Economics,in: Hollis Chenery and T.N. Srinivasan (ed.), Handbook of Development Economics, edition 1, volume 3, chapter 41, pages 2551-2657 Elsevier.

Little, R. J., and Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). Wiley.

Lohr, S. L. (2009). Sampling: design and analysis. Nelson Education.

Lu C, Chin B, Li G, and Murray CJ. (2009) Limitations of methods for measuring out-of-pocket and catastrophic private health expenditures. Bull World Health Organ;87(3):238-44, 244A-244D.

MacDonald, L., Macpherson, D. A., Sirmans, G. S., and Zietz, E. N. (2006). The value of housing characteristics: a meta analysis. The Journal of Real Estate Finance and Economics, 33(3), 215-240.

Malpezzi, S. (2002). "Housing". In Grosh, M. and Glewwe, P. (eds.). Designing Household Questionnaires for Developing Countries, Lessons from 15 years of Living Standards Measurement Study, Volume One: World Bank.

Meyer, B., and Sullivan, J. (2003) "Measuring the Well-Being of the Poor Using Income and Consumption", The Journal of Human Resources, 38.

Meyer, B. D., and Sullivan, J. (2009). "Five decades of consumption and income poverty". National Bureau of Economic Research.

Meyer, B. D., and Sullivan, J. (2011). "Further results on measuring the well-being of the poor using income and consumption." Canadian Journal of Economics, 44(1), 52-87.

OECD (2013). OECD Guidelines for Micro Statistics on Household Wealth.

Oseni, G., Durazo, J., and McGee, K. (2017). The Use of Non-Standard Units for the Collection of Food Quantity. LSMS guidebook.

Oseni, G., Huebler, F., McGee, K., Amankwah, A., Legault, E., and Rakotonarivo, A. (2018), Measuring Household Expenditure on Education: a new guidebook on measurement. LSMS guidebook.

Pape, Utz Johann and Mistiaen, Johan A.. (2018). Household expenditure and poverty measures in 60 minutes : a new approach with results from Mogadishu.

Pradhan, M. (2009). Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia. Fiscal Studies, 30(3-4), 391-417.

Pyatt, G. (1976). On the interpretation and disaggregation of Gini coefficients. The Economic Journal, 86(342), 243-255.

Ravallion M. (1994). Poverty Comparisons.

Ravallion M. (2008) Poverty Lines. In: Durlauf S.N., Blume L.E. (eds) The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

Ravallion, M. (2016). The Economics of Poverty History, Measurement, and Policy. Oxford: Oxford University Press.

Ravallion, M. and B. Bidani (1994). How Robust is a Poverty Profile?, World Bank Economic Review, 8: 75-102.

Rousseeuw, P. J., and Croux, C. (1993). Alternatives to the median absolute deviation. Journal of the American Statistical Association, 88(424), 1273-1283.

Schwabish, J. A. (2014). An economist's guide to visualizing data. Journal of Economic Perspectives, 28(1), 209-34.

Sen, A. (1976). Poverty: An Ordinal Approach to Measurement, Econometrica, 44(2): 219-31.

Sen, A. (1987). The Standard of Living. Cambridge: Cambridge University Press.

Sen, A. (1987). Commodities and Capabilities. New Delhi: Oxford University Press.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. Econometrica: Journal of the Econometric Society, 613-625.

Smith, L. C. (2015). The great Indian calorie debate: Explaining rising undernourishment during India's rapid economic growth. Food Policy, 50, 53-67.

Smith, L. C., Dupriez, O., and Troubat, N. (2014). Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys. International Household Survey Network.

Stigler, G. J. (1954). The early history of empirical studies of consumer behavior. Journal of Political Economy, 62(2), 95-113.

Stiglitz, J. E., Sen, A., and Fitoussi, J. P. (2009). Measurement of economic performance and social progress.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. Public Opinion Quarterly, 74,413–432.

Troubat, N. and Grünberger, K. (2017). Impact of survey design in the estimation of habitual food consumption. The case of the 2007/08 Socio Economic Survey of Mongolia applied to urban households. Food Policy, 72(C): 132–145.

UNECE (2009). Making Data Meaningful, Part 1: A Guide to writing stories about numbers. United Nations, Geneva.

UNECE (2009). Making Data Meaningful, Part 2: A Guide to presenting statistics. United Nations, Geneva.

World Bank Group. (2015). A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals. Policy Research Report. Washington, DC: World Bank.

World Bank. (2017). Monitoring Global Poverty: Report of the Commission on Global Poverty. Washington, DC: World Bank.

Xu K, Ravndal F, Evans DB, and Carrin G. (2009), Assessing the reliability of household expenditure data: results of the World Health Survey. Health Policy;91(3):297-305.

Zheng, B. (1997). Aggregate Poverty Measures, Journal of Economic Surveys, 11(2): 123-62.