

**Preliminary and incomplete: 02/11/2010**

**Standardization of Measurement:  
What Can We Learn from the Economic Sciences?**

Robert J. Willis  
University of Michigan

Prepared for Workshop on  
“Advancing Social Science Theory: The Importance of Common Metrics”

Division of Behavioral and Social Sciences and Education  
Committee on Common Metrics in the Social Sciences  
National Academy of Sciences  
Washington, D.C.  
February 27-28, 2010

## 1. Introduction: Uses of Measurement in Economics and in the Economy

Standardization of measurement is an issue in every science. In economics, standardization is a central aspect of economic behavior as well as a key component of economic science. In this paper, I discuss the role of measurement in economic life and in economic science with a primary aim of suggesting the relevance, benefits and pitfalls of standardization of measurement in economics to other social science disciplines.

## 2. Historical and Political Considerations

Perhaps because economics is so directly relevant to policy and politics in a democratic society, the development of standardized economic data has gone hand in hand with the development of the idea of data in the public service. The founding in 1920 of the National Bureau of Economic Research, the institutional home of the early development of national income accounting and business cycle research, arose from discussions between two men on opposite sides of the political debate in a hearing on minimum wage legislation in New York in 1916. (Fabricant, 1984). These two men, Malcolm Rorty, a politically conservative engineer, and Nahum Stone, an economist with socialist leanings came to agree that "...whenever possible, social programs should rest on objective knowledge of fact and not on subjective impressions, and that the range of such possibilities could be enlarged by scientific research. They knew well enough that facts alone, however firm, could not settle all their own differences on policy... But at least the differences might be narrowed; thinking in public discussions might be clearer; and the discussions might take place on a 'higher level.'" (Fabricant, p. 3) The NBER incorporated these principles into its charter at the time it was founded.<sup>1</sup>

The NBER launched a pioneering series of studies of national income by its founding director of research, Wesley Clair Mitchell, during the 1920s. In 1930, he asked his student, Simon Kuznets, to lead the NBER's national income project and, shortly thereafter, this work was formally transferred to the newly formed Bureau of Economic Analysis of the Department of Commerce under his direction. The creation of national accounts data by professionalized government agencies has since spread across the world. And, the further refinement and extension of the national accounts data has continued to benefit from the interaction between these government professionals and economists located in academics, business and other units of government.

The

---

<sup>1</sup> The NBER was founded with five precepts: (1) Its research should concentrate on determining facts, and the major connections among facts, that are important in dealing with major problems of economic policy. (2) The knowledge sought should be quantitative in character, whenever possible. (3) The research should be in accordance with scientific principles. (4) The research should be done, and the findings made known under auspices and with safeguards that would assure the public of their impartiality. (5) To this end, in particular, the Bureau should carefully abstain from making recommendations on policy (Fabricant, pp. 2-3.)

The NBER founders were far from the first Americans to argue that a well functioning democratic system needs credible data and that credible data cannot be produced if the producers have a stake in what the data show. For example, Kenneth Prewitt (1987) notes that the Bureau of Labor Statistics was established during the period of intense civil strife following the Civil War with the expectation by labor reformers that the presentation of the facts about wages and working conditions to the public and legislators would create a demand for change. He goes on to quote advice given by Francis Walker—the Superintendent of the 1870 Census, the founding BLS Commissioner, the inaugural President of the American Economic Association and a Vice President of the National Academy of Science—to the first Commissioner of the Massachusetts Bureau of Labor Statistics (Walker, 1877: vii-viii):

Your office has only to prove itself superior to partisan dictation and to the seductions of theory, in order to command the cordial support of the press and the body of citizens...I have strong hopes that you will distinctively and decisively disconnect [the bureau] from politics.

The production of valid, credible data requires more than keeping politics at arm's length, as Prewitt (1987, p.118 ) explains,

But even if we accept that professional control over national statistics can largely eliminate fraud and greatly lessen bias in the most important of our social and economic indicators, other issues remain. The statistics of even the most professional agencies suffer from measurement problems for which there are no presently available solutions. When these problems lead to errors of serious magnitude and yet the numbers are used by political leaders to set policy and by citizens to evaluate these policies, the accountability process is compromised.

Alas, as I shall argue in much of the rest of this paper, it has proven necessary for the established statistical agencies to succumb to the seductions of economic theory in order to produce economic data that are useful for science and for policy.

### 3. Measurement in Economic Life

Much of economic life is made up of exchanges between individuals, households or firms. In an exploration of “Measurement Costs and the Organization of Markets,” Barzel (1982, p. 27) begins with the observation that,

“People will exchange only if they perceive what they get to be more valuable than what they give. To form such perceptions, the attributes of the traded items have to be measured. Some measurements are easy to obtain; others pose difficulties”

Much of the textbook literature on the construction of standardized economic measures implicitly assumes that measurements are easy to obtain or, if not easy, that it is feasible

to obtain empirical measures of economic quantities that satisfy the precepts of economic measurement through the use of conventional surveys or administrative data. The “easy case” occurs when information about quantities and prices of the goods that are exchanged is complete in the sense that the attributes of the traded items and terms on which they are traded are measured in the same way by the buyer, the seller and by third parties.

Not surprisingly, measurement problems that “pose difficulties” abound in the real world. They affect researchers, government regulators and policy makers, but most importantly they affect the individuals, households and firms who are engaged in the economic life of a society. Standardized measures in terms of units of measurement of physical quantities or monetary units have ancient origins as do the actions of private actors and sovereigns to subvert the standards for their own advantage. The used car salesman repairs and polishes the parts of the car that are most easily observed by a customer; the quantitative targets of a Soviet planner are optimized by the glass producer in terms of the variables that are rewarded—square feet of glass or tons of glass—not the features that are valued by the consumer—durability and transparency; the troubled bank resists “mark to market” accounting rules and the impecunious mortgage borrower welcomes the opportunity to state his income in order to get a “liars loan;” the doctor prescribes a costly treatment that may have little effect on his patient’s health.

In this paper, I begin by considering the measurement of GDP by methods which, although they have evolved and grown in sophistication since days of Kuznets and the NBER, continue to follow the approach of measuring economic welfare based on objective data on consumption and productive inputs. As the economy has grown richer, the composition of consumption has shifted from easily measurable food and manufactured items with limited variation in quality and variety to one dominated by difficult to measure services and goods with a great range of quality. In addition, economists recognize that the traditional definition of income and consumption to goods and services that are purchased and sold in markets should be extended to encompass the value of leisure and non-market household production, the value of one’s health, the value of public goods, the services of the environment, the value of civil liberties and other factors that influence well being but are not captured in conventional economic measures.

Neuroscientists stress that the brain is an evaluative organ capable of making rapid judgments about which option is better or worse even when differences between the options are complex, ambiguous and difficult to describe in objective terms. Survey researchers, psychologists and sociologists have a long history of attempting to elicit subjective measures of well being by asking questions that elicit subjective responses. Despite their long insistence on the primacy of objective measures, economists are becoming increasingly open to the possibility that subjective questions may offer a viable alternative or complement to objective measures. I will give a few examples of the use of subjective measures in economics later in the paper. While they are promising, these measures are very far from attaining any sort of standardization that commands wide consensus in the economics profession. Research in this area is exciting and, I think,

promises to make economic measurement a more interdisciplinary endeavor that has been in the past.

#### 4. Measurement of GDP

The measurement of prices and output in national income accounts provides the canonical example of standardization of measurement in economic science. In the United States the Bureau of Economic Analysis of the Department of Commerce produces estimates of Gross Domestic Product (GDP using double-entry accounting to measure the monetary value of all the goods and services produced in the economy and the distribution of incomes that production generates. Accounting identities force the total income and total expenditure to be equal. The basis for these aggregate measures lies in micro-level surveys of households, firms, and units of government and administrative records. Similar procedures are used by official government statistical agencies throughout the world.

GDP measures are intended to allow for comparisons of the level of real income across time in a given country and across countries at a given time. In order to carry out these comparisons, it is necessary to adjust for differences in the purchasing power of a monetary unit of income using price indices. A recent National Research Council Report, *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*, provides a masterly and detailed discussion of the connection between economic theory and practical measurement issues (NRC, 2002). Another influential NRC report by Citro and Michael (1995) discusses the measurement of poverty. Both of these NRC reports emphasize the U.S. context. There has also been a major effort carried out over several decades and embodied in the Penn World Tables, to create purchasing power parity (PPP) indices that allow incomes in different countries to be converted into comparable measures (Heston and Summers, 1988; Deaton and Heston, 2010).

Economic theory provides the underpinning for the accounting magnitudes reported in GDP measures. In particular, theories of individual and household behavior and theories of market structure play central roles. To the extent that the structure of the economy corresponds to the theoretical model and the empirical magnitudes that can be measured correspond to the theoretically relevant variables, real GDP measures have a number of important strengths. First, household level real incomes are cardinal measures that allow meaningful mathematical operations to be performed. For example, GDP is the sum of household incomes across the population and this sum may be divided by population to obtain per capita income. Second, income and prices are key determinants of the supply and demand for goods and services by households and firms so that the collection of data on these variables enables the scientific study of economic behavior. Third, real income and related measures provide meaningful, interpersonally and intertemporally comparable measures of welfare that, to an approximation, are not dependent on strong assumptions about individual preferences. Quantiles and moments of the distribution of real income can be compared across subgroups. Thus, these standardized measures provide a basis for policy evaluations at many different levels of aggregation ranging from micro-level studies of individual households or firms, to specific markets, to cities,

regions, nations, and even global studies. Likewise, real income and related measures enable scientific study of economic secular economic growth and business cycles.

## 5. Limitations of Data or Theory?

Despite these immense strengths, the standardized measures used by economics suffer from a number of shortcomings, some technical and of minor importance and others far more serious leading to suggested remedies that contradict longstanding precepts of economic theory. A good example of the latter point is contained in Angus Deaton's 2010 Presidential Address to the American Economic Association (Deaton, 2010). He tells a frustrating tale of a largely failed attempt to determine trends in world poverty measured by the number or proportion of people living on less than \$1 day using cutting edge theory, data and empirical methods to develop a valid PPP measure of the amount of local currency needed to buy a dollar's worth of goods in countries around the world. He ends up presenting evidence from survey questions on self-rated well being from the Gallup World Poll and makes the suggestion that these subjective measures may provide a more accurate picture of trends in world poverty than the more traditional measures.

Deaton's advocacy of using a subjective measure of poverty in place of one based on objective data from consumer expenditure surveys will strike many economists as deeply heretical, a throwback to 19<sup>th</sup> century utilitarianism that had been discarded as unnecessary and unfounded by 20<sup>th</sup> century theorists. The modern economic theory of consumer behavior began with the Marginal Revolution of the 1870s which posited that individuals made their consumption and labor-leisure choices so as to maximize their utility. Until the 1930s, utility was widely regarded as, in principle, a measurable quantity that could be compared across individuals.<sup>2</sup> This view was largely abandoned by economists following the Ordinalist Revolution of Lionel Robbins (1932) and John Hicks and R. G. D. Allen (1934)—codified as “Revealed Preference” by Paul Samuelson (1938, 1947). In mainstream economics, the current definition of utility is simply: *a representation of an individual's preferences over alternatives.*

The problem that Deaton confronts, in my reading of the tale he tells, is not so much the inadequacy of the mainstream theory as it is the difficulty of measuring the variables that the go into this theory. In particular, the methodology underlying the construction of the PPP requires pricing of equivalent baskets of goods in different countries. The problem is households across the globe do not purchase the same goods and, therefore, their consumption baskets cannot be compared. A simple example that Deaton uses to illustrate this problem is how to price men's shirts in New York, London and Nairobi. People in Nairobi do not wear the type of shirt used as the standard in consumer pricing surveys in the U.S. or U.K. If sold in Nairobi at all, such shirts may only be available in hotels catering to foreigners. Their price is irrelevant to poor people in Kenya.

The broader point to make is that the conventional methodology for calculating real income and related measures is strong when data exist on the prices and quantities of all

---

<sup>2</sup> Stigler (1950) provides a classic history of the development of utility theory. Collander (2007) contains a historical discussion focused on the measurability of utility.

of the goods and services that enter into an individual's preferences. In this case, the method of revealed preference may be used to derive variables with the desirable measurement properties that I described earlier. That is, these variables are cardinally measured and interpersonally comparable; they measure important determinants of individual economic behavior and well-being at the micro level; their distributions can be compared in subpopulations and meaningful summary variables may be created from aggregates, moments or quantiles of these distributions such as GDP, per capita income or percentage of the population below a poverty line. However, when the data fail to meet these criteria, the conventional methodology will yield measures of more doubtful quality and much more caution should be used in making inferences from them.

Modern economics has greatly expanded the scope of what it considers to be contained in the set of goods and services that enter a person's preferences. Correspondingly, this has vastly increased the data requirements for the "ideal" data set needed to implement the conventional methodology, often well beyond what could conceivably be collected in surveys or administrative data. Clever use of economic theory has enabled economists to make significant progress in the absence of ideal data. While theory (sometimes clever and sometimes not) has traditionally been the weapon of choice for economists to address data deficiencies, economists are increasingly involved in the design of new data collections using surveys, administrative data and other means to expand and enrich the kinds of data they can use. Economists are also becoming more open to the use of new types of data that fall outside the usually disciplinary boundaries, as illustrated by Deaton's use of self-rated well being data.

In the next section, I present a textbook example to explain how and why the method of revealed preference produces measures with the desirable properties claimed in my earlier discussion when complete data on quantities and prices are available. I then consider a number of examples of ways in which the growing scope of measurement produces challenges standardization of measurement. I conclude by discussing recent development of measures that fall outside the conventional accounting framework used in economics.

## 6. Comparing Apples and Oranges: Revealed Preference and Real Income

Textbooks in economics often present a stylized two-by-two model of a pure exchange economy containing two individuals—say, George and Sarah—who value two goods—say, apples and oranges. Initially, assume that George possesses 20 apples and Sarah possesses 10 oranges. Who is better off? In order to answer this question, it is obviously necessary to compare apples and oranges. It may also seem necessary to compare the values of George and Sarah or perhaps also the values of the third party—the researcher, policy maker or philosopher—who raises the question about comparing the welfare of George and Sarah. By the middle of the 20<sup>th</sup> century, Paul Samuelson's theory of revealed preference (Samuelson, 1938) demonstrated that it is possible to compare the value of 20 apples and 10 oranges and answer the question of whether George or Sarah is better off with only a minimal knowledge of George and Sarah's values and without any need to consult the values of researchers, policy makers or philosophers.

To illustrate the basic ideas underlying the revealed preference approach and explain how it enables economists to provide standardized measures to compare apples and oranges, I return to George and Sarah in a stylized exchange economy. If these two people lived on separate islands, with George living on his endowment of apples and Sarah on her endowment of oranges, it would be impossible to know who was better off or to compare the value of an apple to that of an orange without some measure of welfare or value supplied by a third party such as a psychologist or philosopher or, perhaps, with a direct report from George and Sarah about their well-being that we, as third parties, trust to be interpersonally comparable.

But now suppose that trade is possible. Given that George and Sarah have preferences that value both apples and oranges, George may be willing to give up one of his apples for five oranges while Sarah may be willing to give up one of her oranges for four apples. Clearly, both could gain from trade. In a competitive market, each participant may trade the market price of oranges, expressed as the number of apples that must be given up to obtain an orange. For ease of exposition, let the price of apples be set arbitrarily at \$1. George and Sarah's demand for oranges and apples depend on their preferences and resources, given by their endowments in this example. Competitive equilibrium when the market price of oranges reaches a point at which the quantity of oranges and apples offered for sale is just equal to the quantity demanded of each fruit. At this point, the marginal value of oranges relative to apples to each consumer is equal to their relative price and, therefore, is equated across consumers.

Suppose that the equilibrium occurs when the price of oranges equals \$3. At this price, assume that George consumes 5 apples and 5 oranges by selling 15 apples in order to earn the \$15 income needed to purchase 5 oranges. Similarly, assume that Sarah consumes 7 oranges and 9 apples, using her income of \$9 from the sale of 3 oranges to purchase the apples. Note that Sarah's market income is lower than George's. It would be wrong to conclude, however, that Sarah's economic welfare is worse than George's. Under the weak assumption that George and Sarah both prefer more to less of a given good, the theory of revealed preference implies that Sarah is unambiguously better off than George.<sup>3</sup> Moreover, standard accounting methods used to calculate real income generate the same conclusion.

To show this, note that George can purchase any consumption bundle that costs no more the \$20 cost of the bundle of 5 apples and 5 oranges that he is observed to choose, where cost is evaluated using market prices. Among all such bundles his choice reveals that he prefers (5,5). In a hypothetical experiment, we could offer George his own preferred bundle of 5 apples and 5 oranges or the alternative of any bundle that costs no more the \$30 cost of Sarah's bundle of 15 apples and 3 oranges. Clearly, under the "more is better

---

<sup>3</sup> This axiom is easily extended to cover the case in which an item is a "bad," such as garbage, or even when a given item is a good to some people and a bad to others. For example, I am allergic to peanuts so that, unlike most people, I have strong preference not to consume peanuts. The method of revealed preference can determine whether a given item is a good or a bad in the preferences of a given person simply observing the choice a larger or smaller amount of the item in question.

axiom, he would choose the alternative because, with a \$30 budget he could increase his consumption of apples, oranges or both, thus revealing that Sarah's economic situation is better than his. Conversely, if offered same of her (15, 3) bundle or one costing no more than George's (5,5) bundle, Sarah would reject George's bundle in favor of her own, thus revealing that she also regards her economic situation as better than his.

As a short hand for the rather tedious argument in the preceding paragraph, we commonly say that Sarah's *real income* is higher than George's because it has unambiguously greater purchasing power. Real income can be calculated as the market value of each person's consumption bundle. Alternatively, it can be calculated as the market value of each person's resources, evaluated using market prices, yielding real incomes of \$30 and \$20, respectively, for Sarah and George. The reason that a comparison of their market incomes generates a misleading welfare comparison is that it fails to include the value of non-market consumption of the 5 apples George chooses to consume rather than sell and of 7 oranges that Sarah consumes rather than sells. When the opportunity costs of their non-market consumption is evaluated at market prices and added to market income, the same totals of \$30 and \$20 appear.

This example illustrates how revealed preference methodology allows us to construct a measure of real income that allows interpersonal welfare comparisons to be made without assuming that utility is measurable and with only very weak assumptions about individual preferences. The reason this was possible is that we presumed (a) that the economy is perfectly competitive so that each individual may buy or sell every good that enters into their preferences at common set of market prices and (b) that we have complete data on the quantities of goods that each person consumes and the prices they paid.

To see the crucial role that market prices play, imagine that we observe the quantities of goods consumed by George and Sarah, but have no data on prices. We observe that George consumes more oranges than Sarah, but fewer apples. If the price of oranges is greater than \$5, George would have been able to choose Sarah's bundle and, thus, would have a higher real income than Sarah and conversely if the price of oranges is less than \$5. Without knowledge of how they value apples and oranges there is no way to compare their levels of welfare by revealed preference.<sup>4</sup>

## 7. Index Numbers [to be completed]

Monetary values in the economy may change because of inflation, because of a reallocation of resources between market and non-market production, because of changes in productivity due to technological change, because of business cycle changes and so on. Typically, all of these processes are operating together. A vast literature, beginning in the

---

<sup>4</sup>In the example, we know that George's income comes exclusively from the sale of apples and Sarah's from the sale of oranges. Knowledge of their endowments of apples and oranges together with their consumption bundles would enable us to solve algebraically for the relative price of oranges and apples. Since the link between the prices of the resources that households use to generate income are not closely linked to the prices they pay for the goods and services they consume in the real world, I ignore this avenue for determining prices in the example.

18<sup>th</sup> century (Chance, 1966), has developed procedures for measuring and interpreting economic data that distinguish real and nominal values with applications to the measurement of consumption, earnings and income, productive inputs of labor and capital, rates of interest and so on. Most of these procedures involve the construction of index numbers.

As I mentioned earlier, the NAS has recently published an authoritative volume dealing with index numbers, so I will only briefly discuss them here. [to be done later, include quality variation and tie back to Deaton Presidential talk.]

#### 8. Expanding Scope of Economics to Subjective Measures [to be completed]

The standardized data and index numbers produced by the National Income and Product Accounts correspond to a theoretical framework in which competitive markets exist for the goods and services that enter into household preferences. This framework has been extended and adapted to deal with a variety of complications such as incomplete data, allowing for new products, quality variation in existing products, heterogeneity in consumer preferences and variation in prices due to imperfect. In addition, there has been vast amount of work done on the measurement of capital and labor inputs and productivity change that I will not attempt to describe here.

Despite this progress, growth in the scope of economic theory has outpaced the capacity of the national accounts to produce standardized data

#### 9. Conclusion: Challenges for the Future [to be completed]

References:

Colander, David.(2007) “Edgeworth’s Hedonimeter and the Quest to Measure Utility” *Journal of Economic Perspectives* 21(2): 215–225

Fabricant, Solomon, 1984, Toward a Firmer Basis of Economic Policy: the Founding of the National Bureau of Economic Research  
<http://www.nber.org/nberhistory/sfabricantrev.pdf>

Johnson, David B. (1973), "Meade, Bees and Externalities", *The Journal of Law and Economics* **16** (1): 35–52

Kenneth Prewitt, 1987 "Public Statistics and Democratic Politics," in William Alonso and Paul Starr, eds., *The Politics of Numbers*. New York: Russell Sage Foundation, 1987. 113-128

Samuelson, Paul A. (1938). *A Note on the Pure Theory of Consumers' Behaviour*. *Economica* 5:61-71.

Summers, Robert and Heston, Alan (1988). A new set of international comparisons of real product and price levels estimates for 130 countries, 1950-1985, *Review of Income and Wealth*, 34(1), March, 1-25.