



Claude Hillinger:

Measurement in Economics and Social Science

Munich Discussion Paper No. 2007-19

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at <http://epub.ub.uni-muenchen.de/1928/>

MEASUREMENT IN ECONOMICS AND SOCIAL SCIENCE

Claude Hillinger*

JEL Classification: B16, C43, C82, D61, D71
April, 2007

Key words: collective choice, consumer surplus, money metric, price and quantity indexes, welfare measurement

*SEMECON
University of Munich
Ludwigstr. 33/IV
D-80539 Munich
Germany
E-mail: hillinger@lmu.de

ABSTRACT

The paper discusses measurement, primarily in economics, from both analytical and historical perspectives. The historical section traces the commitment to ordinalism on the part of economic theorists from the doctrinal disputes between classical economics and marginalism, through the struggle of orthodox economics against socialism down to the cold-war alliance between mathematical social science and anti-communist ideology. In economics the commitment to ordinalism led to the separation of theory from the quantitative measures that are computed in practice: price and quantity indexes, consumer surplus and real national product. The commitment to ordinality entered political science, via Arrow's 'impossibility theorem', effectively merging it with economics, and ensuring its sterility. How can a field that has as its central result the impossibility of democracy contribute to the design of democratic institutions?

The analytical part of the paper deals with the quantitative measures mentioned above. I begin with the conceptual clarification that what these measures try to achieve is a restoration of the money metric that is lost when prices are variable. I conclude that there is only one measure that can be embedded in a satisfactory economic theory, free from unreasonable restrictions. It is the Törnqvist index as an approximation to its theoretical counterpart the Divisia index.

The statistical agencies have at various times produced different measures for real national product and its components, as well as related concepts. I argue that all of these are flawed and that a single deflator should be used for the aggregate and the components. Ideally this should be a chained Törnqvist price index defined on aggregate consumption.

The social sciences are split. The economic approach is abstract, focused on the assumption of rational and informed behavior, and tends to the political right. The sociological approach is empirical, stresses the non-rational aspects of human behavior and tends to the political left. I argue that the split is due to the fact that the empirical and theoretical traditions were never joined in the social sciences as they were in the natural sciences. I also argue that measurement can potentially help in healing this split.

1. INTRODUCTION

The story of measurement in economics, including its more recent extension into political science, is one of the strangest, as well as least understood, in the annals of science. It is a story of the entanglement of views about measurement with ideological attitudes of various kinds. In economics, this led theorists to question the relevance of the money metric, the most elementary economic measuring rod, and ultimately to the separation of theory from applications and to the relegation of the subject to statistical agencies. The consequence has been that economists are unaware of the properties, including serious defects, of the data they use.

The rejection of the money metric was generalized in the modern theories of welfare economics and of collective choice, where it was assumed that cardinal representations of preferences are to be ruled out at both the individuals and aggregate levels. On that basis, the economist Kenneth Arrow proved his famous ‘impossibility theorem’¹ that is generally interpreted as meaning that there exists no rational method for aggregating individual preferences; hence, no rational justification of democracy. In Hillinger (2005) I have argued that, since Arrow’s theorem relies on the ideologically motivated rejection of cardinal measurement, it proves no such thing. What the theorem did lead to was the death of political science as a positive discipline able to contribute to the growth of democratic institutions.

In a series of past papers I have dealt with issues of measurement. The measurement of real expenditure and inflation was the subject of Hillinger (2001, 2002, 2003a, 2003b). In Hillinger (2005) I extended the analysis to the representation and aggregation of voter preferences. Each of these articles had an analytical core, but also featured a conceptual inquiry that defined the problem and analyzed the shortcomings of the major proposed solutions. In this context I also looked at the historical background and at ideological precommitments that have blocked the advance of the subject. The conflict between empiricism and various ideological commitments in the history of economics is analyzed at some length in Hillinger (2006). The present paper gives a broad overview of this work. I have also been able to improve the argumentation at some crucial points.

The paper has the following structure: The following section gives a historical overview of the empirical tradition in economics, particularly in relation to measurement. This background is required in order to understand economists’ failure to solve basic problems of measurement and the subsequent abandonment of the subject by academic economists. In Section 3 I clarify basic conceptual issues related to the measurement of real expenditure and inflation. This is followed in Section 4 by a discussion of the principal theories that have attempted to deal with these issues.

The core analytical part of the paper begins in Section 5. I argue that there is a uniquely best method for restoring the money metric and that it involves Divisia integrals and their approximation by means of Törnqvist indexes. I show that the Törnqvist index approximates the Divisia index quadratically for any monotone path of prices and quantities and argue that this solves the issue of path dependence. I relate this result also to an alternative derivation of the Törnqvist index due to Balk and Diewert. Beyond assuming competitive markets, this section makes no specific assumption about how market quantities were generated.

Section 6 considers the case of a utility maximizing consumer and shows that his real consumption and cost of living can be computed in the same manner as before by means of Divisia and Törnqvist indexes. I show that the assumption of homotheticity that has dominated the literature is superfluous. Related results of Theil and Diewert are also discussed.

¹ This is how the theorem is usually referred to. Arrow himself called it the ‘general possibility theorem’.

Section 7 begins the discussion of aggregation, by showing how a Divisia index can be computed as a weighted average of sectoral indexes, the weights being the market shares. The sub-indexes may refer to sectors, such as those of a national economy, or to agents such as consumers. Section 8 defines the aggregate cost-of-living and aggregate real consumption. Section 9 shows how Törnqvist indexes can be computed from sub-indexes, a result of considerable practical value.

The following three sections deal with the computation of the inflation rate and real expenditures in the context of the NIPAs. Section 9 is preliminary in this regard and discusses chain indexes, relating them to Divisia indexes. In Section 10 I argue that real GDP and its components should all be deflated by the same deflator and that ideally this deflator should be a chained Törnqvist price index on aggregate consumption. In Section 11 I propose an entirely new set of accounts that would describe the evolution of *relative prices* of the sectors.

In the concluding Section 13 I discuss the potential of measurement for the unification of the social sciences; specifically the joining of the empirical-sociological with the theoretical-economic. I discuss the entry of happiness research into economics as a step in this direction.

2. A SHORT HISTORY OF ECONOMIC MEASUREMENT IN RELATION TO EMPIRICISM, FORMALISM AND IDEOLOGY¹

2.1. The Empirical Tradition at the Birth of Economics

It is no exaggeration to say that modern economics begins with measurement. This statement refers to William Petty and the ‘political arithmetic’ he created. Before Petty, mercantilist thought was focused on ways to augment the treasure of the sovereign. Petty changed that by defining the total output, or equivalently income, of the economy as the magnitude of central interest. But, he went much further and actually constructed the first measurements of national income and wealth for Ireland and England.

Petty was one of three founding fathers of modern economics, the other two being François Quesnay and Adam Smith. It is useful to consider and to contrast these three in order to understand the origins and characteristics of the empirical tradition and of the not so empirical tradition that came to dominate academic economics. The extensive writings of the three authors on what may be termed social philosophy are not the subject here. Instead, I focus on the one idea that each author is primarily identified with and that has retained its relevance to the present.

Petty was a typical universal genius of the Enlightenment era. Rising from a humble background he became at a young age a doctor in physics, a professor of anatomy and a professor of ‘music’, a term designating at that time a wide field of cultural studies. Drawn to practical pursuits, Petty soon terminated his academic career and enlisted as a physician with Cromwell’s army that invaded Ireland. He was given the task of surveying the land and assessing the riches, some of which were to be distributed among the victors. As a consequence of these endeavors he published in 1691 *The Political Anatomy of Ireland*. In this and in subsequent publications he estimated population, national income and national wealth for Ireland and England. He is thus the father of quantitative economics. The title of his last, posthumously published book *Discourse on Political Arithmetic* contained the title he gave to this new subject of study.

¹ The factual background for this section can be found in any standard history of economic thought; I particularly like Spiegel (1971). Much relevant material can also be found in *The New Palgrave*. Factual aspects of the historical sources are not in dispute, but the interpretation is my own and is quite non-standard. The typical (non-Marxian) history of economic thought tends to be rather uncritical and assumes that the ‘progress’ of economic thought is indeed just that. I view the history as a conflict between empiricism and ideology, with the former often on the losing side. A fuller exposition of this thesis along with references to some recent relevant literature is Hillinger (2006).

Petty, a founding member of the *Royal Society* was personally involved in the scientific revolution occurring in his day. He believed in the methodological unity of the natural and the social sciences and took them to rest on a foundation of measurement. In the introduction to the *Discourse* he expressed his famous credo:

The Method I take to do this, is not yet very usual; for instead of using only comparative and superlative Words, and intellectual Arguments, I have taken the course (as a Specimen of the Political Arithmetick I have long aimed at) to express my self in Terms of Number, Weight, or Measure; to use only Arguments of Sense, and to consider only such Causes, as have visible Foundations in Nature; leaving those that depend upon the mutable Minds, Opinions, Appetites, and Passions of particular Men, to the Consideration of others.

Petty's work inspired some further work in political arithmetic particularly by Charles Davenant and Gregory King.

Quesnay was like Petty a physician, in fact the court physician to Louis XV. Under the label of physiocracy he established a school of social and economic thought. Much of physiocratic thought is antiquated, particularly the idea that all wealth originates in agriculture. The idea that Quesnay is famous for and that has retained its relevance is that of a circular flow of goods and services through the sectors of an economy that could be measured in terms of the associated money flows in the opposite direction; an idea that suggested itself to the physician in analogy to the circulation of the blood. He gave a numerical example of such a system in his *Tableau Economique* (1759), a forerunner of the modern national income and product accounts.

With the work of Petty and Quesnay the basic ideas on which quantitative macroeconomics is based were in place. Thereafter the subject languished until the early Twentieth Century when the rise of socialism and of the econometric movement generated a strong interest in social and economic statistics. The modern equivalents of Petty's and Quesnay's measures, the national income and product accounts (NIPA) are generally regarded as the most important of all economic statistics and enter the national economic and political debates on a daily basis.

Adam Smith, like the other two authors, fully participated in the spirit of the Enlightenment and of the beginning scientific and industrial revolution. The structure of his two principal works was inspired by Newton's theory of universal gravitation. Just as Newton had postulated a single force to explain, among other phenomena, the motions of the heavenly bodies, so Smith postulated in each of his works a single force to explain an aspect of human behavior. In the 1759 *Theory of Moral Sentiments* this force was sympathy; derived from an empathic concern for the feelings of others. In the 1776 *An Inquiry into the Nature and Causes of the Wealth of Nations* he postulated self interest as the driving economic force.

Smith differs from Petty and Quesnay in having never been active in a natural science. Instead, his principal occupation was as a professor of moral philosophy. Nor did Smith share the interest in measurement and quantitative analysis of the other two. Famously he declared "I have no great faith in political arithmetic". Undoubtedly he was expressing a valid skepticism relative to the very rough and uncertain estimates of the political arithmeticians; mere skepticism however does not help a field to advance.

The academic economic mainstream, as it evolved subsequent to Smith, remained essentially in the mold of philosophy, reasoning from intuitively perceived first principles. This was true of classical economics, of the marginalists, of the French mathematical school and, as I shall argue, also of contemporary economics. It is instructive to contrast the beginning of economics with the beginning of the natural sciences. The universities of the early Renaissance, dominated by their theological faculties, were hostile to the rise of science and particularly to the experimentalists who generally came from the class of artisans of lower social standing. The early organization of science took place in scientific organizations outside the universities. It was only when the prestige of science began to replace that of religion that scientists established themselves at the universities, with experimentalists and

theorists enjoying equal prestige. The marriage of empiricism with speculative theory, so characteristic of natural science, never occurred in economics or later social science.

2.2 Philosophical Measurement: Utilitarianism

The very peculiar aspects of economists' relationship to measurement begins with Jeremy Bentham's utilitarianism that became the generally accepted philosophical background of classical economics. Writings on various aspects of utilitarianism could fill a small library; my purpose here is solely to bring out the decisive influence that utilitarianism has had on subsequent attitudes towards measurement.

The Utilitarians postulated that:

- a. There exists a psychological magnitude that they termed 'utility', essentially corresponding to the ordinary language term 'happiness', that it is distinct from money income. While the Utilitarians did not find a way to actually measure utility, they thought that in principle it was measurable on a cardinal scale.
- b. The aim of social policy should be the achievement of 'the greatest happiness of the greatest number'. Formally, they defined this as the sum of individual utilities.

The central tenet of utilitarianism, that it is the purpose of economic activity to maximize the welfare of the general population, as measured by the subjective evaluations of the individuals involved, has remained virtually unchallenged, certainly in the economic mainstream. Where the mainstream departed from the utilitarian position was on the issue of cardinal measurability. I have argued in all of my papers dealing with measurement, either in an economic or in a voting context, that the mainstream's rejection of cardinality was ideologically motivated and had disastrous consequences for the evolution of economics and politics as empirical sciences. The following section summarizes this argument.

2.3 The Rejection of Cardinality.

The marginalists, in their endeavor to replace classical economics, rejected the utilitarian analytical apparatus. They showed that the consumer's optimum is defined by the condition that the rates of substitution between commodities that keep utility constant is equal to the inverse of the ratio of their prices. In this formulation there is no need to consider any utility increments. They argued further that such increments could not be observed and that they were therefore a useless metaphysical construct. This pointed the way towards utility as an ordinal concept, with indifference curves indicating rates of substitution, and utility levels regarded as irrelevant.

This argumentation on the part of the early marginalists was not entirely fair, since they were able to do without cardinal utility only because they ignored the problem for the solution of which cardinal utility was invented: the determination of a social optimum. With the rise of general equilibrium theory this problem came again to the fore. It was solved in the spirit of marginalism by Pareto. He showed that the equilibrium that results under perfect competition has the property that not all individuals can be made better off. This condition of 'Pareto optimality' also makes no use of any utility increments. The Pareto optimum demonstrates the efficiency of a competitive economy. It is weaker than the *optimum optimorum* envisaged by the utilitarians, since the income distribution generated by a competitive economy may be unacceptable. On this the Pareto principle is silent.

The next crucial development was the publication in 1932 of Lionel Robbins' *An Essay on the Nature and Significance of Economic Science*. The essay was directed against the socialist claim that economic and social policy could and should be based on science. Robbins' argument was based on the marginalist position according to which interpersonal utility comparisons are impossible. Actually, Robbins made the more modest claim that such comparisons lacked a scientific basis. Since the determination of a social optimum apparently

requires such comparisons, it follows that policy aiming at the attainment of such a state cannot be entirely scientific.

What Robbins actually said should not have been very controversial. A reasoned reply came from Myrdal (1969) who argued that those who propose a policy should make their values explicit. Unfortunately, Robbins' essay did not lead to a reasoned debate among economists. Instead, the stronger assertion that 'interpersonal comparisons of utility are impossible!' usually preceded by 'Of course!' became a mantra by means of which economists signaled that they were members in good standing of the academic economic mainstream. This strong form of impossibility assumption became the foundation of Arrow's work and of modern collective choice theory generally. This will be discussed further in Section 13.

The conviction of the impossibility of interpersonal utility comparisons also impacted economic measurement. The pervasive economic measure is money. The Utilitarians had made the distinction between money and utility and had given priority to the latter. This amounted to a de-emphasis of the money measure. As already described, the first half of the Twentieth Century saw a great expansion of social and economic statistics, produced with the aim of promoting a scientific approach to the solution of social problems. Many of these measures were expressed in monetary form, most importantly GDP. For economic theorists arose the problem of relating the monetary measures to changes in welfare that in turn had to be defined in relation to utility.

For the GDP, the solution was sought via the Pareto criterion and various compensation measures. It was argued that an increase of GDP signified that the winners could compensate the losers, so that, at least in principle, the Pareto criterion would signify an improvement. The other measure that theorists paid much attention too was consumer surplus, proposed by Dupuit. It was typically used to calculate if the benefit of a project that would result in a lower price to consumers for some service would yield a greater benefit than the cost involved. Around 1950, Paul Samuelson demonstrated that all of these attempts were irreparably flawed. The consequence was that economic theorists largely abandoned problems of measurement. This did not in any way reduce their practical importance; instead, theory and practice parted ways.¹

It is ironic that cardinal measurement was abandoned by theorists in economics and political science, precisely the two fields that have natural cardinal measures in the form of money and the ballot. The opposite path was taken in psychology. Here central concepts such as intelligence, or various emotions, have no natural metric, but immense efforts were made to construct such metrics. Sophisticated statistical methods such as factor analysis and multi-dimensional scaling were developed for the purpose.

As already described, my own work on measurement and on voting essentially involved the rejection of the self-defeating and ideologically motivated commitment to ordinalism; instead, I elaborated the theory of the naturally available cardinal measures.

3. REAL EXPENDITURE AND INFLATION: A CLARIFICATION OF BASIC CONCEPTS AND TERMINOLOGY

My aim is to give a unified and satisfactory treatment of a subject that has been treated in an unsatisfactory and fragmented manner. For the purpose of orientation I give a list of terms that have been used to designate one or the other of the two sides of the problem: price index, quantity index, inflation, real expenditure, consumer surplus, consumer price index, real national product and its components, economic growth. This list is not exhaustive. For example, NIPA statisticians have adopted the term 'volume index' instead of 'quantity index',

¹ More detail and references to the literature can be found in my papers dealing with the money metric cited above.

apparently realizing that a quantity index is not analogous to a quantity; of course, it is even less analogous to a volume.¹

The problem discussed under these diverse headings is essentially always the same: to divide a change in total expenditure into one part that is due to the change in quantities and another that is due the change in prices. The basis for this interpretation is simple. The measures fall into two categories: those based on changes, in either difference or ratio form, of prices and those based on changes of quantities. The price changes have quantities as weights and the quantity changes have prices as weights. When prices and quantities are both variable, the two kinds of measures divide the total expenditure change into one due to the changes in quantities at fixed price weights and one due to the changes in prices at fixed quantity weights. The measures differ in the algebraic formulas that are used and in the weights that are chosen. When the weights are prices, either the initial price vector, or the final price vector, or some combination is used; similarly when the weights are quantities.

While there are thus two kinds of measures, there is only one aim: to restore the money metric that is lost when prices vary. This is done either directly by means of a quantity index, or indirectly by deflating the nominal expenditure with a price index.²

A clear and appropriate terminology usually goes hand in hand with clear and appropriate concepts. The proliferation of inappropriate terms in economic measurement is indicative of the confusion of concepts that in turn is related to the separation of theory from applications discussed in the previous section. All of the measures discussed here involve products of prices and quantities and are thus expenditures; as is appropriate if they are to be used to decompose changes in nominal expenditure. For the expenditure from which the effects of price changes have been removed the term *real expenditure* is well established and will be used here. This term can be used regardless of whether the change is reported as a difference or as a ratio. No specific terminology exists to denote the complimentary concept of an expenditure from which the effects of quantity changes have been removed. In fact, unlike the case of real expenditure, where it is important to have an estimate of the *level*, there is no need for a comparable expenditure level that depends on price changes only. Where the complementary concept is important is in relation to changes, and most importantly in relation to proportional changes. The reason for this is that proportional changes are independent of the units of measurement. For example, considering two economies of very different size and with different monetary units, the comparison of real expenditures, or price changes, in absolute units makes no sense; as proportional rates of change such comparisons are common place. For the proportional change in expenditure that is due to price changes I will use the term *inflation*. This requires an explanation since inflation is defined as the average proportional rate of increase of prices. However, when quantities are constant and all prices change in the same proportion, expenditure changes in that proportion also. This is one justification for the use of the term. A more rigorous justification comes from the theory of this paper. The measure for the proportional change in expenditure due to price changes is the Törnqvist price index, which is a geometric average of the proportional price changes.

Having defined what is to be measured in principle, I turn to the logically prior question of why? This issue is little discussed in the literature because of the rejection of the money metric by economic theorists. The measuring rod provided by money is constantly employed by individuals in their economic calculations. In the absence of precise measures, people attempt an intuitive adjustment for the changing price level. To make these calculations explicit and as accurate as possible is the task of the relevant economic theory.

¹ Lequiller and Blades (2006 p.20) give the following justification: "...national accountants prefer the term "GDP in volume" to "real GDP" because inflation is just as real as growth."

² McKenzie (1983) also takes the position that the money metric is the central theoretical concept relevant to welfare measurement.

The usual terminology when referring to measures in ratio form is ‘price index’, ‘quantity index’, or ‘index number’ to refer to either type. This terminology is illogical and reflects a misunderstanding of what is actually being measured. Specifically, in theories using the concept of a representative agent, price and quantity indexes have been used as inputs to the representative agent’s decision making in the same manner as ordinary prices and quantities are used as inputs to the decision making of normal agents. There is no justification for this practice. This topic is discussed further below. Apart from this, the term ‘index number’ is nonsensical. A formula is not a number and an index is a symbol used for counting. In this paper I will heed the injunction of Occam’s razor, not to unnecessarily multiply concepts, and will use established terminology where it would be inconvenient to change it; for example, I will refer to a ‘Törnqvist quantity index’.

4 CURRENT THEORIES

4.1 The Axiomatic Approach

This goes back to the *test approach* of Irving Fisher. He postulated certain properties that an index should have and used these to derive the index that bears his name. The axiomatic approach, as it has evolved, suffers from the following drawbacks:

- a. More axioms have been proposed than can simultaneously be satisfied. Depending on which subset of axioms is chosen, different formulas result.
- b. The source and hence the validity of the axioms has not been made clear.
- c. The axiomatic approach does not clarify the economic implications of the numbers computed using a given index.

4.2 Economic Theories

Theories covered under this heading are all based on the assumption of a maximizing agent. The agent could be a utility maximizing consumer, or a profit maximizing firm. Most of the literature deals with the consumer’s cost-of-living index. The duality between inflation and real expenditure indexes plays hardly any role in this theory. In fact, the quantity index most often discussed is the Malmquist index, which is not based on the money metric and is not dual to an inflation index.

A further assumption that is made is that the ‘aggregation function’ $f(\mathbf{x})$, which may be the utility function of a consumer, or the production function of a firm, is linear homogeneous, a rather unrealistic assumption in the case of the consumer. Based on these assumptions, the economic theory divides into two major branches.

One branch is the theory of ‘superlative indexes’ due to Erwin Diewert. He has shown that there exists a class of indexes that approximate the true change in inflation, or real expenditure. In the case of the consumer these would be the changes in the cost-of-living and in real consumption. The approximation is quadratic *in the limit* as the increments of prices and quantities go to zero. The Fisher and Törnqvist indexes are the most prominent members of the class of superlative indexes.¹

The other branch is the *econometric index theory*. This approach is explicitly based on the concept of a representative agent, usually a representative consumer. A specific parametrized representation of the direct or indirect utility function is assumed, for example a homogeneous quadratic function in the logarithms of prices or quantities. The corresponding demand functions are derived and are estimated econometrically from market data. From the estimated

¹ Diewert has published extensively and over a long period of time. A wide ranging and up to date account is Diewert (2001).

parameters of the demand functions, those of the utility function are computed; these in turn allow the computation of the changes in real consumption and the cost of living.¹

The econometric approach has been the major occupation of theorist working in this area for several decades, but it has not been adopted by the statistical agencies. The immediate reason is undoubtedly the very high level of complexity and sophistication that is required. More fundamental is the fact that the approach is flawed because of its reliance on the representative agent concept. Since this concept is central to contemporary macroeconomics, I give a brief separate discussion.

4.3 *The Representative Agent*²

The representative agent is a very strange concept because it has been rejected by virtually everyone who has seriously considered the question of its justification; this rejection however has done nothing to curtail its popularity. The representative agent model would be justified if it were possible, starting from the assumption of individually maximizing agents, to deduce that observable aggregate variables would behave as though they were the outcome of the maximizing behavior of a single agent. That such a drastic reduction in complexity should be possible is implausible on the face of it. Aggregation theory confirms this judgment and shows that the result is possible only under implausible conditions. Hartley (1977) writes:

Under what conditions will we be able to derive a consistent representative agent (or equivalently a macro model)? The most important conditions are...If all agents have parallel, linear Engel curves, or equivalently, if all agents have identical homothetic preferences, consistent aggregation is possible...

Since Gorman wrote, there have been several papers providing additional conditions which will yield consistent aggregation... There is no real need for us to explore all of these conditions in detail. It is enough to note that every one of them is thoroughly implausible; it would be remarkable in and of itself if anyone argued that any of these functional forms was in any way realistic. Indeed, Lewbel (1989, p. 631), after an entire paper devoted to developing general forms which allow for aggregation, concludes, "It is a fact that the, use of a representative consumer assumption in most macro work is an illegitimate method of ignoring valid aggregation concerns." (p. 134).

In concluding his review of the literature on representative agents Kirman (1992) writes:

Given the arguments presented here – that well-behaved individuals need not produce a well-behaved representative agent; that the reaction of a representative agent to change need not reflect how the individuals of the economy would respond to change; that the preferences of a representative agent over choices may be diametrically opposed to those of society as a whole – it is clear that the representative agent should have no future. (p. 134).

Franklin M. Fisher (1987) gives a broad review of all aspects of the aggregation problem; over consumers, over firms, and over commodities. He concludes:

Such results show that the analytic use of such aggregates as 'capital', 'output', 'labour' or 'investment' as though the production side of the economy could be treated as a single firm is without sound foundation. This has not discouraged macroeconomists from continuing to work in such terms.

The conclusions of the theory of representative agents are devastating, but only half of the story, since the theory deals only with aggregation over agents. Empirical applications require equally aggregation over commodities. For example, the representative firm requires inputs of 'labor' and 'capital' for which quantity indexes are used. The representative consumer uses quantity indexes representing 'housing', 'food', etc. There is no justification for this kind of use of quantity indexes even in the case of a concrete agent.³

Both kinds of aggregation are involved in the econometric theory. The theory of superlative indexes, when applied to market data, involves aggregation over agents, but not over commodities. The axiomatic theory does not use the representative agent concept, but

¹ A survey of the econometric approach is Slesnick (1998). The most ambitious realization of the approach is Jorgensen (1990).

² Hillinger (2006) discusses this topic at some length and gives further references to the literature.

³ A valid form of aggregation over commodities is 'Hicksian aggregation' involving a group of commodities for all of which prices change in the same proportion. The limited usefulness of this concept is evident.

neither does it solve the problem for which that concept was invented: to enable inferences from aggregate data.

4.4 Consumer surplus

Consumer surplus is one of the most important concepts in economics. Two facts may be adduced in support of this assertion. It has generated a huge literature to which many prominent economists contributed.¹ It is widely used throughout the world as a tool for the evaluation of public projects. Given the importance of the concept, size of the relevant literature and the scientific renown of many contributors one would expect that some clarity and consensus would have been achieved. The following quotations cited in Hillinger (2001) show that this is not the case.

...to this day consumers' surplus discussions involve ambiguities, confusions, approximations, redundancies and Napoleonic pretensions hard to match in other parts of economic theory. (Samuelson, 1990, p. 275)

The concept of consumer surplus is extremely useful in making many decisions about public goods – it has been employed in decisions about airports, roads, dams, subways, and parks. (Samuelson and Nordhaus, 1985, p. 418)

... $1/2(p_0 - p_1)(q_1 - q_0)$...is a formula used over and over again in cost – benefit analysis, especially for a small change in prices so the linearity assumption is a reasonable approximation to any actual demand curve.

(Layard and Glaister, 1994, p. 50)

Despite its ubiquity it is now widely accepted that Consumer's Surplus should not be used as a welfare measure...

(Slesnick, 1998, p. 2108)

Particularly noteworthy is the fact that Samuelson, who as a theorist consistently rejected consumer surplus, along with other cardinal measures, switches to an entirely positive evaluation in the many editions of his elementary text, now jointly co-authored with Nordhaus.

In the following I will try both to give a reasoned evaluation of the concept and to throw some light on its confused history.

The fascination that the concept has exerted on generations of economists is undoubtedly related to the following: **a.** It is derived from the most fundamental of all economic concepts that of a utility maximizing consumer and his derived demand curve. **b.** The measure has a simple, intuitively satisfying geometric derivation that has been a staple of introductory and intermediate textbooks. **c.** The formula is simple and additive so that aggregate consumer surplus can easily be computed from market data.

Assume for some commodity a price decline and accompanying increase in the quantity purchased. Assuming also a linear demand function, the monetary value attached by the consumer to the additional units purchased is given by the 'welfare triangle' $1/2(p^0 - p^1)(x^0 - x^1)$. To this must be added the amount saved on the purchase of the initial quantity given by $(p^0 - p^1)x^0$. Adding the two terms gives

$$(4.1) \quad CS^1 = \frac{1}{2}(x^0 + x^1)(p^0 - p^1).$$

The argument so far corresponds to that originally advanced by Dupuit (1844); as the above quotation from Layard and Glaister shows it is still the standard argument advanced in applications oriented texts. A conceptual problem with Dupuit's argument was pointed out by Marshall²: Implicit in the argument is the assumption that the marginal utility of income is constant as one moves along a consumer's demand curve. This issue has bedeviled the subsequent literature. Another problem is a proliferation of both terminology and alternative

¹ Takayama (1987) reviews the literature.

² Marshall's position is discussed by McKenzie (1983, Ch. 4).

formulas that have all been used in connection with consumer surplus. Thus (4.1) is actually a measure of the change in the consumers cost-of-living. The measure that is usually computed for his change in welfare, the change in real expenditure in the terminology of this paper, is

$$(4.2) \quad CS^2 = \frac{1}{2}(p^0 + p^1)(x^1 - x^0).$$

Alternatively, $-CS^1$ has also been used as the welfare measure, since for $\Delta y = 0$, $-CS^1 = CS^2$.

There have been many attempts at a formal derivation of consumer surplus measures. These have involved more general measures that allow all quantities and prices to be variable. I will refer to the generalized expression of the form (4.1) as a *centralized price variation* (CPV) and to the generalized equivalent to (4.2) as a *centralized quantity variation* (CQV)

$$(4.3) \quad CPV = \frac{1}{2}(\mathbf{x}^0 + \mathbf{x}^1)(\mathbf{p}^1 - \mathbf{p}^0), \quad CQV = \frac{1}{2}(\mathbf{p}^0 + \mathbf{p}^1)(\mathbf{x}^1 - \mathbf{x}^0).$$

These expressions were introduced into the literature by Bennet (1920). He argued similarly to Dupuit that CPV represents areas under demand curves and that CQV represents areas under inverse demand functions. This argument is even more dubious in this general case since, when all prices and money incomes are variable, demand curves between a single good and its price are not defined.

Three authors are cited as having given rigorous derivations of CQV as a measure of individual or aggregate welfare changes. They are Hotelling (1938), Hicks (1941-42, 1945-46) and Harberger (1971). If one scrutinizes their derivations carefully, one finds in each case a step that depends on the implicit assumption of a constant marginal utility of income. Samuelson (1942) pointed out that given that the indirect utility function is homogeneous of degree one in \mathbf{p}, y , the marginal utility of income cannot be constant.

At this point I make a detour from discussing the evolution of ideas in the general economic literature to the evolution of my own ideas on the subject. I believe that this 'internal history' can shed considerable light on both the nature of the problem and on its collective history. Interest in economic measurement goes back to the beginning of my career as an economist since I was convinced that the connection between measurement and theory was essential for a meaningful empirical science. At the same time I saw the complete separation of theory from measurement in economics, related in particular to Samuelson's uncompromising criticism of cardinal measures, particularly consumer surplus and real national income. I felt this to be an intolerable situation.

The two measures were connected in my mind because of the additivity property of CS; a property that I felt had to be shared by real GDP and its components. Connected with this belief was an implicit assumption that still seems to be held by NIPA statisticians – the components of the real NIPAs are to be computed by means of quantity indexes and should add to total GDP given by its own quantity index. It required a long intellectual evolution on my part to rid myself of this assumption.

Initially I felt that the problem was to obtain a rigorous derivation of CS on the basis of which I could then advocate its use in the computation of real GDP. I looked at this problem intermittently over the years, but without success. The breakthrough came when I saw the article by Diewert (1992). It contained an improved exposition of Hick's attempted derivation as well of the precise reason for its failure. Based on Diewert's exposition I was able to see at once how the proof could be fixed. This was the first mathematically correct proof that given a well behaved, but otherwise arbitrary indirect utility function, a quadratic approximation to the consumer's real expenditure change is given by CQV and to his cost-of-living is given by CPV. Having obtained the first proof, I found many more. The final result reported in Hillinger (2001) was that there exist an infinity of plausible theoretical definitions for both the change in real consumption and the cost-of-living all of which are either approximated

quadratically by the CS measures, or are even reproduced exactly. At the time I was rather proud of this abundance of results. Later I came to view it rather as inelegance inherent in the discrete approach. I now believe that it is better to avoid this ambiguity by taking limits and first looking at the situation at a point, as is done in the Divisia approach.

Given the importance that has been attached to CS for nearly two centuries and given the many unsuccessful attempts by prominent economists at a rigorous derivation, one might have expected the above results to make an impact. This did not happen; the SSCI does not record a single citation to the article. The reason appears to be that theorists have abandoned the subject of measurement while practitioners are more interested in an intuitive graphic demonstration than in mathematical rigor.

Having given a rigorous derivation of the CS formulas I investigated the possibility of their use in computing real GDP and its components. The problem that immediately arises is that the formula

$$(4.4) \quad \Delta \mathbf{x}^t = 1/2(\mathbf{p}^{t-1} + \mathbf{p}^t)(\mathbf{x}^t - \mathbf{x}^{t-1})$$

cannot be used to extrapolate real GDP and its components over several periods since each increment would be valued at a different average price. To use the formula at all deflated prices have to be used. In Hillinger (2002) I derived a suitable deflator for this purpose and advocated the use of a formula analogous to (4.4), but with deflated prices. However, as Ehemann et al. (2002) pointed out, this can lead to the computation of a negative value for a real component. This led me to abandon the proposal and to conclude that the only reasonable method for computing real GDP and its components is to use a single deflator. This is discussed further in Section (11).

4.5 The Divisia/Törnqvist Approach¹

This is the approach that in my view can provide a satisfactory theoretical foundation for the measurement of real expenditure and inflation in a variety of contexts. The past elaboration of this theory has not been satisfactory, accounting for the limited attention that it has received. The Divisia/Törnqvist approach can also be regarded as the natural science inspired approach to the problem. In the natural sciences a problem is usually analyzed first at a point of time using differentiation. The solution over time is then derived by solving the resulting differential equation. Similarly, Divisia approached the problem of separating the effects of price and quantity changes on expenditure by first taking partial derivatives of expenditure with respect to these variables. The solution between two points of time could then be obtained by integration. Divisia pointed out that the standard discrete approximation of his differential leads to a Laspeyres index if the initial slope is used and to a Paasche index if the final slope is used. These indexes are linear approximations.

The fundamental problem that is connected with Divisia's idea is that the integral is path dependent. A literature exists that derives exact solutions for some highly artificial time paths of prices and quantities, but a practical implication of these results is not apparent. The best known idea for an approximation is that of Törnqvist who noticed that the integral has an exact solution when market shares are constant. The precise relevance of this result has also remained unclear. Törnqvist did not prove any approximation property of his index to the general case of variable shares. moreover, any estimate of the value of the integral is of doubtful value unless one finds a way of defusing the path dependence property. If, for given endpoints, the value of the integral can be made arbitrarily large in both the positive and negative directions, the value of any particular estimate must be questioned.

In Hillinger (2003a) I made a decisive advance in the justification of the Divisia/Törnqvist approach. I proved that the Törnqvist index is a quadratic approximation to the Divisia integral for *any monotonic path*. Combine this result with the economic insight that only the

¹ Balk (2000) has provided a general review of work on Divisia indexes. My own contribution on which much of the next section is based is Hillinger (2003b).

endpoints are actually relevant for the comparison between them. This makes any oscillating path between the endpoints irrelevant. Since all monotonic paths agree to a quadratic approximation, it is reasonable to take this common value as the difference between the endpoints.

The approximation theorem just described is quite general, making use only of the definition of expenditure as the product of prices and quantities. The power of the Divisia integral does not end here. I showed that it works equally well in measuring real expenditure and inflation in more specialized contexts. In particular, it can be applied to the case of a single maximizing agent, particularly the consumer. This application does not involve the usual restriction to a homogeneous aggregator function. The theory even works, without any special restrictions, for an aggregate of agents, such as consumers. Unlike the usual invocations of a representative agent, the result here is completely rigorous.

In the present paper I also describe a second path leading to the Törnqvist index. It combines the Divisia differential with a set of axioms due to Balk and Diewert (2000).

5. MEASURING REAL EXPENDITURE AND INFLATION: THE FOUNDATION

5.1 Bennet and Divisia Differentials

Let $\mathbf{p}^t = (p_1^t, \dots, p_N^t)$, $\mathbf{x}^t = (x_1^t, \dots, x_N^t)$, be a pair of price, quantity vectors. Their vector product¹ gives the total expenditure: $\mathbf{p}^t \mathbf{x}^t = \mathbf{x}^t \mathbf{p}^t = \sum_i p_i^t x_i^t = y^t$. Throughout the paper, competitive prices are assumed in the relevant markets. Bennet (1920) laid the foundation of the continuous approach by assuming that total expenditure could be regarded as the product of a price and a quantity index:²

$$(5.1) \quad P(t)Q(t) = \mathbf{p}(t)\mathbf{x}(t).$$

He also defined the corresponding differentials

$$(5.2) \quad QdP = \mathbf{x}d\mathbf{p}, \quad PdQ = \mathbf{p}d\mathbf{x}.$$

Based on these differentials, Bennet suggested the following discrete approximations to the indexes, which for simplicity I define for the time periods 0, 1:

$$(5.3) \quad P^1 - P^0 = (1/2)(\mathbf{x}^0 + \mathbf{x}^1)(\mathbf{p}^1 - \mathbf{p}^0)$$

$$(5.4) \quad Q^1 - Q^0 = (1/2)(\mathbf{p}^0 + \mathbf{p}^1)(\mathbf{x}^1 - \mathbf{x}^0).$$

It is easy to check that the two expressions are dual, i.e., they add exactly to the total expenditure change.

Divisia converted the Bennet differentials to proportional form, which makes them independent of units of measurement. The Divisia price differential is

$$(5.5) \quad \frac{PQ \frac{dP}{P}}{PQ} = d \ln P = \frac{\sum p_i x_i \frac{dp_i}{p_i}}{y} = \sum s_i d \ln p_i, \quad s_i = \frac{p_i x_i}{y}.$$

Similarly, the Divisia quantity differential is

$$(5.6) \quad d \ln Q = \sum s_i d \ln x_i.$$

The two differentials decompose the change in expenditure:

$$(5.7) \quad d \ln y = d \ln P + d \ln Q.$$

The decomposition has two paramount features:

a. The real growth rate of the aggregate is a weighted average of the component growth rates and the inflation rate is a weighted average of the proportional price changes, the weights being the expenditures shares.

¹ I use a notation that suppresses the transpose sign since it is inconvenient when vectors have superscripts.

² A discussion of the related work of Bennet and Divisia is found in Diewert (1998) and Balk (2000).

b. Real growth and inflation rates are dual so that real growth computed directly, or indirectly via deflation, has the same value.

5.2 Divisia Differentials and the Money Metric

The elementary conception of the money metric is that it is given by money expenditure at constant prices. Actual prices are never constant, yet there are many circumstances in which we still wish to use the money metric; there is therefore a need for a suitable generalization. I believe that the most natural generalization is provided by the Divisia differential and by implication also by the Divisia integral and its Törnqvist approximation. At any point, the differential (5.6) gives the increase in expenditure that is due to the quantity changes only. The magnitude of the computed change in real expenditure is independent of the price *level*, but not independent of *relative prices*. Consider a given quantity change from an initial vector \mathbf{x}^0 to a final vector \mathbf{x}^1 alternatively at one of the price vectors \mathbf{p}^* or \mathbf{p}^{**} . Assume that for those quantities that are increasing the corresponding prices are higher in \mathbf{p}^{**} and for those that are declining the prices are higher in \mathbf{p}^* . Clearly, the measured change in real expenditure will be larger under \mathbf{p}^{**} than under \mathbf{p}^* . This effect is direct in the case of the Bennett differential; in the Divisia differential, the effect occurs through corresponding changes in the value shares.

That the money metric is not invariant to relative price changes makes economic sense and is related to the old utilitarian insight that subjective utility is not the same as money income or expenditure. If the prices of goods on which we would like to spend additional income are very high, the marginal utility of such income is low.¹

The position taken in this paper is that it is the task of the theoretician to provide the best possible measure of the money metric and at the same time to make the limitations of the measure explicit. The task of applied economic policy is to use the measure with discretion and an awareness of its limitations. This issue will be discussed further in the concluding section.

5.3 Divisia Integrals

The point decomposition (5.7) can only be a starting point, since in an empirical context we will always be interested in comparing two or more distinct observations. A necessary step in that direction is to define the integrals corresponding to the Divisia differentials.

The Divisia price and quantity integral are

$$(5.8) \quad V_p = \ln \frac{P^1}{P^0} = \int_0^1 \sum s_i(\tau) \frac{p_i'(\tau)}{p_i(\tau)} d\tau, \quad p_i'(\tau) = \frac{\delta p_i}{\delta \tau},$$

$$(5.9) \quad V_Q = \ln \frac{Q^1}{Q^0} = \int_0^1 \sum s_i(\tau) \frac{x_i'(\tau)}{x_i(\tau)} d\tau, \quad x_i'(\tau) = \frac{\delta x_i}{\delta \tau}.$$

These integrals are path-dependent. Their sum is the integral of the total differential of logarithmic expenditure and thus path-independent:

$$(5.10) \quad V_p + V_Q = \int_0^1 d \ln y(\tau) = \ln \frac{y^1}{y^0}.$$

The Divisia indexes corresponding to the integrals are

$$(5.11) \quad P_D = \frac{P^1}{P^0} = \exp V_p, \quad Q_D = \frac{Q^1}{Q^0} = \exp V_Q.$$

¹A related phenomenon occurred in the former planned economies. Basic necessities were in relatively good supply, but such goods as automobiles, or better quality housing were severely rationed. Additional income was therefore of limited utility, and the efforts to obtain it were not intense.

In connection with the Divisia integrals three issues need to be dealt with:

- a. How can numerical values for the integrals be computed.
- b. What is the implication of path dependency and how can it be dealt with.
- c. What is the economic interpretation of the integrals?

I will deal with these issues in the same order.

Quadratic Approximation

I give two slightly different proofs of the quadratic approximation property of Törnqvist indexes. The first assumes that all variables grow at constant rates. This is the most reasonable assumption one can make if one assumes a specific path. This path can also be given a normative interpretation: If the actual path is unknown, than the integral should be given the value associated with the most regular path. The second proof only requires the assumption of monotone paths. Both proofs are based on the

*Trapezoid Rule:*¹

$$(5.12) \quad \int_a^b f(x) dx = \frac{b-a}{2} [f(a) + f(b)] - \frac{(b-a)^3}{12} f''(c), \quad c \in [a, b].$$

The first term on the right is the trapezoidal approximation to the area above (or below) the interval $b-a$, based on the height of the function at the endpoints. The second term is the residual, which is cubic in Δx .

The theorem will be proven in relation to the price index, the case of the quantity index being analogous. In order to employ the scalar form of the trapezoid rule we write the i th component as

$$(5.13) \quad V_{iP} = \int_0^1 s_i(\tau) \frac{p_i'(\tau)}{p_i(\tau)} d\tau.$$

First Törnqvist Approximation Theorem: Assume that prices and quantities grow at constant rates. Then

$$(5.14) \quad \exp V_p = \frac{P^1}{P^0} = P_T + O_3, \quad \exp V_Q = \frac{Q^1}{Q^0} = Q_T + O_3.$$

Proof:

Letting r_i be the rate for the i th price, it is determined by

$$(5.15) \quad p_i^1 = p_i^0 \exp r_i, \quad \Rightarrow r_i = \ln \frac{p_i^1}{p_i^0}.$$

Then

$$(5.16) \quad V_p = \int_0^1 \sum s_i(\tau) \ln \frac{p_i^1}{p_i^0} d\tau.$$

The i th component

$$(5.17) \quad V_{iP} = \int_0^1 s_i(\tau) \ln \frac{p_i^1}{p_i^0} d\tau$$

is of the standard form given in (5.12), so that

¹ For a discussion of the rule and related results see Judd (1998, Section 7.1). The trapezoid rule is closely related to the quadratic approximation lemma given in Diewert (1976a) and used there for a different derivation of the Törnqvist index in the context of the economic theory of indexes. For an exhaustive treatment of the lemma and its applications in index theory see Diewert (2000).

$$(5.18) \quad V_{iP} = \bar{s}_i \ln \frac{p_i^1}{p_i^0} + O_3(\Delta\tau), \quad \bar{s}_i = \frac{1}{2}(s_i^0 + s_i^1).$$

As shown above, this approximation leads to the Törnqvist price index.

Putting the results together, we can write the approximation of the theoretical Divisia price index by the Törnqvist price index as

$$(5.19) \quad P_D = \exp V_P = P_T + O_3(\Delta\tau).$$

An analogous derivation for the Törnqvist quantity index gives

$$(5.20) \quad Q_D = \exp V_Q = Q_T + O_3(\Delta\tau).$$

The assumption of constant growth rates is the most natural and simplest assumption that can be made in order to prove the quadratic approximation property of Törnqvist indexes. Nevertheless, it is interesting to ask if the result holds under more general conditions. This is the subject of the

Second Törnqvist Approximation Theorem: Assume that: prices and quantities grow monotonically in the interval (0, 1). Then (5.19) and (5.20) hold.

Proof:

Given the monotonicity assumption, there is a unique correspondence between each price or quantity and the corresponding value share. We can therefore write

$$(5.21) \quad V_{iP} = \int_{\ln p_i^0}^{\ln p_i^1} s_i(\ln p_i) d \ln p_i.$$

This expression is of the form given in (5.12) so that

$$(5.22) \quad V_{iP} = \bar{s}_i (\ln p_i^1 - \ln p_i^0) + O_3(\Delta \ln p_i),$$

which is analogous to (5.18). The implication is that (5.19) and (5.20) hold.

Path Dependence

The issue of path dependence should be looked at not only from a mathematical point of view, but also in the light of additional information provided by economic theory. The basic intuition embodied in most economic theory is that a comparison of expenditures between two points of time depends on those points only, and not on the path between them.¹ An oscillating path, that for given endpoints can produce arbitrary values of the index, has no meaning. Conversely, along monotone paths, all Divisia indexes agree to a quadratic approximation with the Törnqvist index. This common value is the most reasonable estimate of the difference between the two points.

5.4 Axiomatic Derivation and Interpretation of Törnqvist Indexes

In this section I discuss the meaning of measures of real expenditure and inflation and relate this to the axiomatic derivation of index formulas and thus also to an alternative derivation of the Törnqvist index. The axiomatic approach always involves a direct comparison of two price/quantity situations without the intermediate step of differentiating at a point as in the Divisia approach. It will be seen that the two approaches are actually complementary.

The axioms are usually stated for price indexes, though analogous axioms apply equally to quantity indexes. A bilateral price index, i.e., one involving a comparison between two situations, can be defined in either in terms of the price and quantity vectors, $P(\mathbf{p}^0, \mathbf{p}^1, \mathbf{x}^0, \mathbf{x}^1)$, or in terms of the vectors of prices and of expenditure shares

¹ Path dependence of tastes or technology is of course always a possibility, but it is not the usual assumption in the context of economic measurement.

$P(\mathbf{p}^0, \mathbf{p}^1, \mathbf{s}^0, \mathbf{s}^1)$ Using the first definition, the axiomatic approach typically leads to the Fisher index, while the second definition leads to the Törnqvist index.

A particularly concise and elegant derivation of the Törnqvist price index is due to Balk and Diewert (2001). Their derivation is based on three assumptions:

The Index is a Function of Value Shares and Price Ratios:

$$(5.23) \quad \ln P(\mathbf{p}^0, \mathbf{p}^1, \mathbf{s}^0, \mathbf{s}^1) = \sum m_i(s_i^0, s_i^1) \ln \left(\frac{p_i^1}{p_i^0} \right),$$

where $m_i(\)$ is an, as yet unspecified, averaging function. The authors do not explain how they arrived at this formulation so that the reader may suspect that it is simply the assumption that is required to obtain the desired result. A good argument can however be made in relation to the Divisia differential (5.5). We see that (5.23) directly adapts (5.5) to the discrete case. Instead of the instantaneous share, an unspecified average of the end point shares is postulated; integrating the logarithmic price differentials yields the logarithmic price ratios.

The authors further assume two of the most basic axioms for price indexes:

Proportionality in Current Prices:

$$(5.24) \quad P(\mathbf{p}^0, \lambda \mathbf{p}^1, \mathbf{x}^0, \mathbf{x}^1) = \lambda P(\mathbf{p}^0, \mathbf{p}^1, \mathbf{x}^0, \mathbf{x}^1), \text{ for } \lambda > 0.$$

Inverse Proportionality in Base Period Prices:

$$(5.25) \quad P(\lambda \mathbf{p}^0, \mathbf{p}^1, \mathbf{x}^0, \mathbf{x}^1) = \lambda^{-1} P(\mathbf{p}^0, \mathbf{p}^1, \mathbf{x}^0, \mathbf{x}^1), \text{ for } \lambda > 0.$$

These axioms reflect fundamental economic intuitions about properties that a deflator should have. Balk and Diewert show that (5.23), (5.24) and (5.25) imply that

$$(5.26) \quad P = P_T.$$

The authors considered only the derivation of a price index. The Törnqvist quantity index could be derived from analogous axioms applied to the quantity vectors.

Of course, these are not the only axioms that the Törnqvist indexes satisfy. Diewert (2004) lists 17 axioms that the Törnqvist price index satisfies. However, based on the axiomatic approach alone, it cannot be said that the Törnqvist index is best. Diewert (2001) found that the Fisher index satisfies more axioms than the Törnqvist index. An important property that the Fisher index does have, but the Törnqvist index does not, is what I refer to as duality and what is known in the literature as the factor reversal test; it is the property that the product of the price and quantity indexes should give the expenditure ratio for the two periods. Here again the connection with Divisia integrals is important. Divisia indexes have the duality property; since Törnqvist indexes are quadratic approximations to Divisia integrals, they have the duality property to a quadratic approximation. In other words, the deviation from duality is a residual in the third and higher powers of the variables. Given all of the errors that one usually encounters in economic data, such a discrepancy is not very serious.

The decisive superiority of the Törnqvist index over the Fisher index derives from its connection to the Divisia integral that allows an elegant integration into economic theory without requiring the unrealistic assumption of a linear homogeneous aggregator function. This is the subject of the next section.

6. THE RATIONALITY ASSUMPTION

6.1 The Continuous Approach

The existing literature on the application of the Divisia index to the problem of the utility maximizing consumer has focused on the assumption of homotheticity. This leads to an

elegant theory that avoids the path dependency of the usual Divisia index.¹ In this section, I present the Divisia theory for the non-homothetic, but rational consumer (household)².

The following definitions will be used: Let \mathbf{x} be the household consumption vector, \mathbf{p} the corresponding price vector, $y = \mathbf{p}\mathbf{x}$ the household expenditure and $u(\mathbf{x})$ a utility function, assumed twice continuously differentiable and strictly quasi-concave. The corresponding expenditure function

$$(6.1) \quad e(\mathbf{p}, u) = \min_{\mathbf{x}} \mathbf{p}\mathbf{x} : u(\mathbf{x}) \geq u$$

specifies the minimum expenditure required to reach the utility level u at prices \mathbf{p} . The expenditure function is the fundamental tool for aggregating prices and quantities in this context. How this is to be done in the general non-homothetic case has not been clarified in the received theory. I propose to do this analogously to the preceding sections by using continuity in order to arrive at unambiguous parameterizations. I also adopt a terminology appropriate for the consumer sector: the inflation measure will now be referred to as the *cost-of-living* (C) and the real expenditure measure as *real consumption* (R). We now require that

$$(6.2) \quad C(t)R(t) = e(\mathbf{p}(t), u(t)) = y(t).$$

The increment in expenditure due to an increment in the cost-of living is defined as

$$(6.3) \quad RdC = \nabla_{\mathbf{p}} e(\mathbf{p}, u) d\mathbf{p}$$

and the increment in expenditure due to the increment of real consumption by

$$(6.4) \quad CdR = \nabla_{\mathbf{x}} e(\mathbf{p}, u(\mathbf{x})) d\mathbf{x}.$$

These increments decompose the expenditure change so that

$$(6.5) \quad de = dy = RdC + CdR.$$

These results are analogous to those for the Divisia differentials. The difference is that \mathbf{x} is now not arbitrary, but rather the solution to the consumer's maximization problem (6.1). The money metric is now defined in relation to the differential (6.4) and can be alternatively referred to as money metric utility or real consumption.

Further progress requires the following

Lemmas on duality of the expenditure function

Let $\mathbf{h}(\mathbf{p}, u)$ be the Hicksian (compensated) demand function.³

$$(6.6) \text{ (Hotelling)} \quad \nabla_{\mathbf{p}} e(\mathbf{p}, u) = \mathbf{h}(\mathbf{p}, u) = \mathbf{x}.$$

$$(6.7) \text{ (Balk)} \quad \nabla_{\mathbf{x}} e(\mathbf{p}, u(\mathbf{x})) = \mathbf{p}.$$

Where \mathbf{x} must be the solution to (6.1)

Converting (6.3) to logarithmic form and using (6.6) gives

$$(6.8) \quad \frac{RC}{C} \frac{dC}{C} = d \ln C = \frac{\sum p_i \frac{\partial}{\partial p_i} e(\mathbf{p}, u) \frac{dp_i}{p_i}}{e(\mathbf{p}, u)} = \frac{\sum p_i x_i \frac{dp_i}{p_i}}{y} = \sum s_i d \ln p_i.$$

Similarly, using (6.4) and (6.7)

¹ This theory is reviewed in Balk (2000, Section 8) and in Diewert (2001, Section D.1)

² The conditions under which a household, as opposed to an individual consumer, can be assumed to be utility maximizing are the subject of a literature that began with Samuelson (1956) and was elaborated further by Pollak (1980).

³ Hotelling's lemma is standard fare of microeconomic textbooks. For the proof of Balk's lemma see Balk (1989, p. 166).

$$(6.9) \quad \frac{CR \frac{dR}{R}}{CR} = d \ln R = \frac{\sum x_i \frac{\partial}{\partial x_i} e(\mathbf{p}, u) \frac{dx_i}{x_i}}{e(\mathbf{p}, u)} = \frac{\sum x_i p_i \frac{dx_i}{x_i}}{y} = \sum s_i d \ln x_i.$$

The logarithmic differentials of C and R are precisely those obtained earlier in the case of the Divisia inflation and real expenditure integrals. Under the same monotonicity assumption as before, these differentials can be integrated and the integrals approximated quadratically by Törnqvist indexes.

$$(6.10) \quad \frac{C^1}{C^0} = P_T + O_3, \quad \frac{R^1}{R^0} = Q_T + O_3:$$

The interpretation of the Törnqvist index Q_T is now that the proportional increase in money metric utility of the consumer is the same as would have obtained if nominal expenditure had increased in that proportion at constant initial prices. The same interpretation obtains for the indirect measure $\frac{y_1/P_T}{y_0}$.

The interpretation just given is subject to some qualification and is made more precise in the next section. *Money metric utility* is defined by the expenditure function $e(\mathbf{p}^0, u(\mathbf{x}))$, for a given base period price vector \mathbf{p}^0 . From (6.9) it is seen that the Divisia differential for real consumption gives the change in expenditure due to the change in consumption and hence utility *at the instantaneous price*. Integration takes place over a changing money metric. The construction of the Divisia and Törnqvist indexes is such that they are not affected by scale effects and therefore not by changes in the price *level*, but they can be affected by changes of *relative* prices as well as by the utility level. The results of the next section clarify this matter further.

6.2 The Discrete Approach

The fixation of index theory on the assumption of a homogeneous aggregator function is the more surprising as Theil (1967, 1968), in a brilliant but neglected contribution, developed the theory of the general case for the individual utility maximizing consumer. Only his assumptions and results are given here, the reader is referred to the original paper for the proofs.

Theil begins his analysis by defining the theoretical index of the cost-of-living, also known as the Konüs cost-of-living index.

$$(6.11) \quad P_K(\mathbf{p}^1, \mathbf{p}^0; u^*) = \frac{e(\mathbf{p}^1, u^*)}{e(\mathbf{p}^0, u^*)},$$

where the reference utility level u^* remains to be determined.

The real consumption index, also known as the Allen quantity index, is defined as

$$(6.12) \quad Q_A(u^1, u^0; \mathbf{p}^*) = \frac{e(u^1, \mathbf{p}^*)}{e(u^0, \mathbf{p}^*)},$$

with the reference price vector \mathbf{p}^* to be determined. Theil's definition of real consumption is a version of money metric utility normalized by \mathbf{p}^* . He explicitly points out the consequence of non-homotheticity: C is not independent of u^* and R is not independent of \mathbf{p}^* . In order to determine \mathbf{p}^*, u^* he assumes that \mathbf{p}^* is an average of $\mathbf{p}^0, \mathbf{p}^1$ and that u^* is determined by the indirect utility function $u^* = u(y^*, \mathbf{p}^*)$, where y^* is the same average of y^0, y^1 , as \mathbf{p}^* is of $\mathbf{p}^0, \mathbf{p}^1$. There follow five elementary conditions of symmetry and homogeneity for the

averaging function that narrow it down to the geometric one. Specifically, we must have

$$p_i^* = (p_i^0 p_i^1)^{\frac{1}{2}}, \quad y_i^* = (y_i^0 y_i^1)^{\frac{1}{2}}.$$

Having obtained unique expressions for the theoretical indexes, Theil turns to the question of their approximation. I state here only the results:

$$(6.13) \quad P_K = P_T + O_3, \quad Q_A = Q_T + O_3.$$

The theoretical indexes are approximated quadratically by the corresponding Törnqvist indexes.

Diewert (1976a, Theorem 2.16) obtained a similar result for the Törnqvist price index via a different route. He showed that on the assumption that the consumer maximizes a general, quadratic, non-homothetic, translog utility function

$$(6.14) \quad C(\mathbf{p}^1, \mathbf{p}^0; u^*) \equiv P_T, \quad u^* = (u^1 u^0)^{\frac{1}{2}}.$$

The results of this section can be summed up as follows: The change in household expenditure can be decomposed into two parts. One is the change in real consumption, the other the change in the cost of living. The theoretical magnitudes can be defined by means of continuously changing parameters, or by means of discrete parameters that are averages of values taken at the endpoints of the interval. In either case, quadratic approximations are given by the appropriate Törnqvist indexes. It should be mentioned that the continuous theory described in this paper is considerably simpler.

6.3 Homotheticity

The assumption of homothetic preferences has been prominent in theories of welfare measurement and economic theories of index numbers. In the econometric approach to welfare measurement homotheticity enables aggregation over consumers. In the theory of bilateral indexes homotheticity is required in order to obtain ‘invariant’ indexes, which will be defined below. Finally, in the theory of Divisia indexes homotheticity is the condition for path independence. The role of homotheticity in relation to the last two topics is fully explored in Samuelson and Swamy (1974) and Balk (2000). Here I only report the principal results.

Consider again the theoretical indexes

$$(6.15) \quad P_K(\mathbf{p}^1, \mathbf{p}^0; u^*) = \frac{e(\mathbf{p}^1, u^*)}{e(\mathbf{p}^0, u^*)} \quad \text{and} \quad Q_A(u^1, u^0; \mathbf{p}^*) = \frac{e(u^1, \mathbf{p}^*)}{e(u^0, \mathbf{p}^*)}.$$

Invariance means that P_K must be independent of u^* and Q_A must be independent of \mathbf{p}^* . Writing Q_A as a function of the optimizing consumption vectors and imposing the duality condition gives

$$(6.16) \quad \frac{y^1}{y^2} = P_K(\mathbf{p}^1, \mathbf{p}^0) Q_A(u^1, u^0).$$

Samuelson and Swamy prove that (6.16) implies and is implied by the assumption of homothetic preferences.

Samuelson and Swamy also demonstrate that homotheticity is both necessary and sufficient for path independence of the Divisia integrals. Balk proves that under this assumption $P_D = P_K$. From the duality property of these indexes it also follows that $Q_D = Q_A$. The final result that should be mentioned here is that if the consumer’s utility function is Cobb-Douglas, then expenditure shares are constant and the Törnqvist indexes are exact for the Divisia indexes. This was Törnqvist’s original insight.

The received theory evidently considers the properties of invariance and path independence as being of great importance, but it is neither made clear whence this importance derives, nor what should be done, given that homotheticity is not a realistic

assumption. In the introduction to their paper, Samuelson and Swamy state that only the homothetic case allows *all* of Irving Fisher’s conditions to be met; they do not elaborate which conditions are otherwise violated, or why this should be a concern. Following is the final section of their paper, titled *Concluding Warning*:

Empirical experience is abundant that the Santa Claus hypothesis of homotheticity in tastes and in technical change is quite unrealistic. Therefore, we must not be bemused by the undoubted elegances and richness of the homothetic theory. Nor should we shoot the honest theorist who points out to us the unavoidable truth that in nonhomothetic cases of realistic life, one must not expect to be able to make the naive measurements that untutored common sense always longs for; we must accept the sad facts of life, and be grateful for the more complicated procedures economic theory devises.

The authors do derive some rather intricate bounds on the theoretical indexes for the non-homothetic case. These apply only to the single utility maximizing consumer and have had no consequence for applications.

Regarding Divisia integrals and the issue of path independence, Balk (p. 3) summarized the existing literature as follow:

A fundamental property of the Divisia indices, which they share with chain indices, is their so-called path-dependency. Over the years this property has led to conflicting views among economists. On the one hand, by a minority this intriguing, almost “magical”, property was considered as a virtue. It was thought that the Divisia indices somehow track economic reality better than (simple) bilateral indices. On the other hand, quite a number of economists have wrestled with this property as a problem and sought after conditions under which the indices may exhibit path-independency.

This is a description of the very unsatisfactory state of the literature. Apart from the fact that there is no agreement, neither of the two positions offers a solution. The first statement is completely vague; it neither clarifies the nature of the allegedly present additional information, nor offers a means of using it. The second position led to the mathematical results on homotheticity that are not satisfied in the real world.

My views on these issues differ from those just described. My view on path-dependence is implicit in my derivation of the Törnqvist index. There I showed that it is not a problem as long as the paths are monotone, since to a quadratic approximation all such paths are valued by the Törnqvist indexes and oscillating paths can have no relevance for the comparison between the endpoints.

Regarding invariance, my position is that if a measure gives the correct answer in particularly simple cases, then the confidence we attach to that measure in more complex cases increases. I believe that I have presented strong arguments in favor of Törnqvist indexes that are unrelated to homotheticity. The properties under homotheticity are elegant, but not of central importance.

7. AGGREGATION OF DIVISIA INDEXES OVER AGENTS AND SECTORS

Up to this point we considered Divisia and Törnqvist indexes as aggregators of prices and quantities pertaining to a single unit, be it a household or a market. This section considers aggregation over such units. Unless dealing specifically with aggregation over households, I will use the term ‘sector’. The method of aggregation is essentially the same, only that there are now three different kinds of expenditure shares to be considered: The share of the i th good in the k th sector s_{ik} , the share of the i th good in the total s_i , the share of the k th sector’s expenditure in the total σ_k . These are related by

$$(7.1) \quad s_i = \sum_k s_{ik} \sigma_k, \quad i \in (1, \dots, I), \quad k \in (1, \dots, K).$$

The logarithmic Divisia price index for the aggregate is

$$\begin{aligned}
(7.1) \quad V_P &= \ln \frac{P^1}{P^0} = \int_0^1 \sum_i s_i(\tau) \frac{p_i'(\tau)}{p_i(\tau)} d\tau \\
&= \int_0^1 \sum_k \sigma_k(\tau) \sum_i s_{ik}(\tau) \frac{p_i'(\tau)}{p_i(\tau)} d\tau \\
&= \sum_k \int_0^1 \sigma_k(\tau) \sum_i s_{ik}(\tau) \frac{p_i'(\tau)}{p_i(\tau)} d\tau.
\end{aligned}$$

Denoting by P_k the Divisia price index for the k th sector, we can also write

$$(7.2) \quad V_P = \sum_k \int_0^1 \sigma_k(\tau) d \ln P_k(\tau).$$

Similarly,

$$(7.3) \quad V_Q = \sum_k \int_0^1 \sigma_k(\tau) d \ln Q_k(\tau).$$

The aggregate integral is a weighted average of the instantaneous Divisia differentials, the weights being the instantaneous market shares. This is analogous to how the price or quantity changes are weighted in the single sector Divisia differential.

The aggregation properties of the Divisia index are all that is really needed since they are inherited by the Törnqvist index. Nevertheless, it is interesting to directly derive the corresponding results under the rationality assumption of the preceding section. Also interesting is the direct derivation of the aggregation properties of the Törnqvist index. These are the subjects of the next two sections.

8. DIVISIA AGGREGATION OVER RATIONAL HOUSEHOLDS

The theory for the individual household can be extended to an aggregate of households on the assumption that the market price is the same for all consumers. The k th consumer, $1 \leq k \leq K$, has expenditure y_k and faces market prices \mathbf{p} . The aggregate consumption vector is $\mathbf{x} = \sum \mathbf{x}_k$. The collection of utilities is $\mathbf{u} = (u_1, \dots, u_K)$. Aggregate expenditure is $y = \sum y_k = \mathbf{p} \sum \mathbf{x}_k = \mathbf{p} \mathbf{x}$. Define $\mathbf{X} = (\mathbf{x}_1 \dots \mathbf{x}_K)$ and the aggregate expenditure function $e(\mathbf{p}, \mathbf{u}(\mathbf{X})) = \sum e_k(\mathbf{p}, u(\mathbf{x}_k))$. The gradient of $e(\cdot)$ wrt \mathbf{p} is given by

$$(8.1) \quad \nabla_{\mathbf{p}} e(\mathbf{p}, \mathbf{u}) = \nabla_{\mathbf{p}} \sum e_k(\mathbf{p}, u_j) = \sum \mathbf{x}_k = \mathbf{x}$$

It would be nice if we could have an analogous gradient wrt \mathbf{X} of the form

$$(8.2) \quad \nabla_{\mathbf{x}} e(\mathbf{p}, \mathbf{u}(\mathbf{X})) = \mathbf{p}.$$

This seems at first sight nonsensical since \mathbf{x} is not an argument of $e(\cdot)$. The expression would make sense if we could show that

$$(8.3) \quad \nabla_{\mathbf{x}} e(\mathbf{p}, \mathbf{u}(\mathbf{X})) \Delta \mathbf{x} = \mathbf{p} \Delta \mathbf{x}$$

because (8.2) could then be viewed as an instruction to compute $\Delta e(\cdot)$ according to the formula

$$(8.4) \quad e(\mathbf{p}, \mathbf{u}(\mathbf{X}^1)) - e(\mathbf{p}, \mathbf{u}(\mathbf{X}^0)) = \mathbf{p} \Delta \mathbf{x} + O_2(\Delta \mathbf{X}).$$

The validity of (8.4) follows from

$$(8.5) \quad \sum \nabla_{\mathbf{x}_k} e_k(\mathbf{p}, u_k(\mathbf{x}_k)) \Delta \mathbf{x}_k = \sum \mathbf{p} \Delta \mathbf{x}_k = \mathbf{p} \Delta \mathbf{x}.$$

The derivation is based on Balk's lemma (6.7) and the assumption that all households face the same price vector. The interpretation of (8.2) is that, when the variations of the $\Delta \mathbf{x}_k$ are small and their sum is given, their distribution is immaterial for the determination of $\nabla_{\mathbf{x}} e(\cdot)$. An alternative derivation of (8.2) is to regard it as an implication of (8.1), given duality.

With these preliminaries, we are in a position to define the logarithmic differentials of the *Aggregate Cost-of-Living C* and of *Aggregate Real Consumption R*. Using a vector notation

$$(8.6) \quad d \ln C = \frac{\nabla_{\mathbf{p}} e(\mathbf{p}, \mathbf{u}) d \mathbf{p}}{e(\mathbf{p}, \mathbf{u})} = \frac{\mathbf{x} d \mathbf{p}}{y} = \boldsymbol{\sigma} d \ln \mathbf{p}$$

$$(8.7) \quad d \ln R = \frac{\nabla_{\mathbf{x}} e(\mathbf{p}, \mathbf{u}(\mathbf{X})) d \mathbf{x}}{e(\mathbf{p}, \mathbf{u}(\mathbf{X}))} = \frac{\mathbf{p} d \mathbf{x}}{y} = \boldsymbol{\sigma} d \ln \mathbf{x}.$$

The differentials are those of Divisia integrals, this time defined on the vectors of aggregate consumption quantities and their prices. The appropriate indexes therefore again have the Törnqvist form.

9. AGGREGATION OF TÖRNQVIST INDEXES

The summation of proportional changes along an interval generally requires an integral, since the shares that serve as weights vary continuously. It is a remarkable property of Törnqvist indexes that they can be aggregated exactly, using only the initial and final shares. For this purpose, the Törnqvist price index is written as the product of two geometric price indexes, one involving the initial and the other the final period.

$$(9.1) \quad P_T = \left[\prod_i \left(\frac{P_i^1}{P_i^0} \right)^{s_i^0} \right]^{\frac{1}{2}} \left[\prod_i \left(\frac{P_i^1}{P_i^0} \right)^{s_i^1} \right]^{\frac{1}{2}} \\ = (P_T^0)^{\frac{1}{2}} (P_T^1)^{\frac{1}{2}}.$$

Let K denote the index set of the k th sector. The k th sectoral index is

$$(9.2) \quad P_{Tk} = \left[\prod_{i \in K} \left(\frac{P_i^1}{P_i^0} \right)^{s_{ik}^0} \right]^{\frac{1}{2}} \left[\prod_{i \in K} \left(\frac{P_i^1}{P_i^0} \right)^{s_{ik}^1} \right]^{\frac{1}{2}} \\ = (P_{Tk}^0)^{\frac{1}{2}} (P_{Tk}^1)^{\frac{1}{2}}$$

The aggregating equation is

$$(9.3) \quad P_T = \prod_k \left[(P_{Tk}^0)^{\sigma_k^0} (P_{Tk}^1)^{\sigma_k^1} \right]^{\frac{1}{2}}.$$

Similarly,

$$(9.4) \quad Q_T = \prod_k \left[(Q_{Tk}^0)^{\sigma_k^0} (Q_{Tk}^1)^{\sigma_k^1} \right]^{\frac{1}{2}}.$$

Unlike the literature on approximate aggregation of indexes, this aggregation is exact. In addition to its theoretical interest, it can also be used for efficient computation, for example in the context of the NIPAs. Once a set of indexes have been computed at a given level of aggregation, the raw data used for these computations is no longer required in computing the indexes of the next higher level.

10. CHAINING

Thus far, we analyzed bilateral comparisons based on the implicit assumption that the price and quantity vectors being compared are not too different, so that a reasonable approximation of the empirical to the theoretical measures will result. In a time series context, a bilateral index is suitable for year-to-year comparisons. It has long been recognized that a fixed index base cannot be maintained for too long, because as the changes in the variables become large the accuracy of the quadratic, or any other, approximation declines sharply. The alternative is some form of chaining. It has also been recognized that chaining introduces path dependence, usually referred to as violation of Fisher's circularity axiom. This has left practitioners in a quandary. The past practice in the context of the NIPAs has been to keep the base constant for 5 or 10 years and then to do some kind of rebasing to establish comparability of the different segments. The problems involved in this will be discussed further in the next section. Currently opinion has shifted towards the use of annually chained indexes. Furthermore, the view, at least of theoreticians, is that a symmetric index, not the usual Laspeyres formula should be used.

From a theoretical point of view, a chain index may be regarded as an approximation to a Divisia index over the entire interval. If year-to-year Törnqvist indexes are used, a sequence of quadratic approximations to the continuous path is obtained. From a numerical point of view, using more points of interpolation and thus more information, increases accuracy. In the present context, a limit to this improvement is set by annual data. Quarterly or monthly data introduce additional drift due to seasonal fluctuations. In addition, the accuracy of the data declines sharply. At the other end, the traditional method of holding the base constant over longer periods is pointless. The underlying continuous index is not changed thereby, only the approximation to it is worse.

For completeness, I state here how chain indexes can be used to compute levels. This is done by means of the usual convention that identifies the initial real magnitude with the nominal expenditure. The implied initial price level is 1. Let P^t be the price level and Q^t the real expenditure level, both at time t . In terms of the theoretical indexes

$$(10.1) \quad P^t = (1) \left(\frac{P^1}{P^0} \right) \cdots \left(\frac{P^t}{P^{t-1}} \right), \quad Q^t = (y^0) \left(\frac{Q^1}{Q^0} \right) \cdots \left(\frac{Q^t}{Q^{t-1}} \right)$$

The theory of the present paper suggests that the indexes used in the above expressions should be Törnqvist indexes.

11. THE DEFINITION OF REAL MAGNITUDES IN THE NIPAS

In Hillinger (2003b) I argued that the statisticians who produce the national income and product accounts (NIPA) do not understand how to compute the accounts in real terms. Here I will try to clarify the essence of the problem. Theorists have produced a large number of axioms for bilateral indexes, but curiously none seem to have been proposed for the components of a real expenditure. I propose such an axiom now.

A.1 Axiom for the Components of Real Expenditure: Any two real expenditures that refer to the same price level must be proportional to the corresponding market values.

This axiom reflects the most basic property of markets: things that have the same value can be exchanged.

There is a related axiom that NIPA statisticians have implicitly attempted to follow in the past, but subsequently abandoned. It is:

A2. Axiom on Additivity: Given nominal expenditures that add up to some total, additivity must be maintained when both the components and the total are converted to real expenditures.

The first axiom implies the second, but not *vice versa*.

The background is that the NIPA statisticians presented both the nominal and the real accounts as additively consistent, up to a statistical discrepancy attributed to measurement errors. The problem that they faced was that they employed sector specific quantity indexes for each sector and a global quantity index for the total; additivity is thereby necessarily destroyed.¹ Additivity was then restored by distributing the discrepancy over the components. This evidently arbitrary procedure was followed without publicity and did not come to the attention of economists and econometricians. The sophisticated econometric methods that were employed on these data were essentially meaningless. This is a dramatic example of the fact that one cannot do meaningful empirical science without a thorough knowledge of the data generating process. In a salutary decision, the NIPA statisticians have abandoned these corrections and instead publish the discrepancies. The implication is that there no longer exists an additively consistent set of accounts in real terms.

The question naturally arises if there is any substantive argument in favor of the assumption that real magnitudes for sectors should be computed as sectoral quantity indexes. The NIPA statisticians meet annually in Paris under the auspices of OECD. When, at the September 2000 meeting I argued that GDP and its components should be deflated by means of a single deflator, the response was that this would be unacceptable to the users of the statistics; an evident non-argument.

A response to and rejection of my proposal on the part of NIPA statisticians can be found in Ehemann, et al. (2002, p. 47). Following are their arguments and my responses:

Equation (12), which divides the value of each subaggregate in current prices by a common, aggregate price index, appears to provide data users with very little information beyond what is already provided in the aggregates valued at current prices. Most cross-sectional comparisons focus on shares or relative magnitudes, which are not affected by use of a common deflator. As we argue above, the measures valued at current prices are generally more appropriate for this type of analysis anyway. For econometric investigation of hypotheses provided by economic theory, the success of models that include real variables obtained by deflation by sector-specific price indexes is well established, and we noted above the contribution made by substituting Fisher for fixed-weighted indexes. Econometric testing would quickly reject models based on allegedly “real” magnitudes obtained with a common deflator.

The authors do not state what information they have in mind. The information provided by a series on real expenditures is simply that it gives the nominal expenditure change that is due to changes in quantities, not in prices. The authors go on to praise the methodology that they themselves abandoned and for good reason. What additional information is provided by cooked accounts involving arbitrary transfers between the sectors in order to restore additivity? That econometric models based on these data achieved significant test statistics is not surprising in view of the facts that macroeconomic series are strongly correlated and that there has been much preselection of models, with those not yielding significant statistics remaining unpublished. The fact is that the huge efforts devoted to macroeconometrics have not led to a generally accepted and validated explanation of macroeconomic dynamics.

The new methodology proposed by the authors is to compute annual ‘contributions to growth’ that are additive. They involve using average prices of the two years involved. Since the prices change from year to year, they do not allow the construction of time series giving *levels* of real magnitudes and therefore cannot be used in structural macroeconomic modeling.

I am forced to conclude that the NIPA statisticians have no idea on how to produce a set of accounts in real terms. GDP and its components should all be deflated using a single deflator. The deflator should be a chained Törnqvist index. In Hillinger (2003b) I have also argues that the index should be defined on consumption.

¹ The easiest way to see this is to consider the real magnitudes as having been obtained through multiplication of each nominal magnitude by a deflator. When all the parts of a sum as well as the total are multiplied by different numbers, additivity is clearly lost.

Addendum:

Having essentially completed the present paper, I obtained a copy Lequiller and Blades (2006), the most recent comprehensive OECD publication on the national accounts. In Chapter 2 they discuss the procedures used to create real (in their language ‘volume’) accounts. They are quite critical regarding non-additive sectoral accounts computed by means of chained quantity indexes and state that this practice is followed only by the US and Canada. The methods used by other countries and by OECD itself are described somewhat sketchily. My understanding of their account is that real sector levels are obtained by extrapolating a base year using a chained Laspeyres quantity index. Real GDP is then defined as the sum of the sectors. This procedure has the consequence that the share of the real sectors in the total will drift away from those of the nominal shares. This in turn has means the implied *relative prices* between the real sectors are not the actual relative prices at which market transactions can take place. A model based on such data cannot be an adequate representation of an economy.

12. AN INTEGRATED SYSTEM OF ACCOUNTS FOR MEASURING INFLATION

In Section 9 I discussed the aggregation of Törnqvist indexes over component indexes. Such an aggregation process is particularly interesting for the Törnqvist price indexes and suggests the creation of a system of accounts showing how inflation at each higher level derives from the inflations of the components. At present the public discussion of inflation is concentrated on very few indexes: most importantly the CPI, to a lesser extent the index of producer prices and rarely the GDP deflator. The sectoral determination of these indexes is reported only episodically.

I believe that a set of three such accounts would be most informative. The first would show annual rates of inflation. From this table one could, for example, see how much of the CPI inflation of a given year, or quarter, was caused by each of its components. A second account would present the corresponding price levels, starting from a value of unity in some base year. In a final account, all sectoral price levels would be ‘deflated’ by the general price level. This would be a table of ‘relative prices’. If for some sector k and period t the table shows that $P_k^t = 2$, the implication is that, starting from the base period, prices in that sector increased twice as much as the average for the economy. A system of accounts for relative prices would be a genuine novelty and an increase in economically meaningful information

13. MEASUREMENT AND THE FUTURE OF SOCIAL SCIENCE

Measurement is at the interface of theory and empiricism. A scientific field cannot advance very far on the basis of casual observation; it requires data that have been systematically generated using methods of measurement that are themselves based on existing theory. In the introductory section I argued that that the natural sciences emerged from the merger of an empirical with a theoretical tradition and that this merger was never successfully accomplished in economics; the sad state of the economic theory of measurement is one consequence. In the present section I broaden the discussion to the present state of the social sciences, with emphasis on the aspect of measurement. I also attempt to connect the present state of the social sciences with their beginnings and to sketch the implications for a possible future. This endeavor requires some drastic simplification, but I believe that a valid outline of the story can be told.

The imperial expansion of economics in recent decades has often been remarked upon. It has two aspects: One is that economists now study phenomena previously dealt with in sociology, political science, or social psychology. The other aspect is that scholars in these fields increasingly adopt the methodology of theoretical economics that emphasizes the use of abstract models based on the assumption of rational behavior, particularly game theoretic models. For each of these developments a paradigmatic example can be given. For the

influence of economics on social science I will discuss the impact of Arrow's 'impossibility theorem' on political science. The impact of social science on economics is illustrated by 'happiness research'. These two movements also illustrate two polar opposite styles: The abstract, analytical style of mathematical economics that came to characterize social choice theory and the concrete, empirical style, derived from sociology and psychology that characterizes happiness research. Both fields deal with the same problem: the measurement and aggregation of preferences, or value judgments, but neither makes an impact on the other.

In Hillinger (2006) I dealt extensively with voting theory and the related field of collective choice, specifically with Arrow's theorem. The principal conclusion was that the commitment to an ordinal representation of preferences has blocked the way to the only possible solution, which is to represent preferences on a cardinal scale. The work of a number of other authors has pointed in the same direction. In my paper I dealt at some length with the history of the ideological rejection of cardinality, a history also touched upon in the introductory section of the present paper.

My work on voting has focused on the assumptions of ordinality versus cardinality since this choice has decisive implications for the theory. The ordinality assumption, so dominant in contemporary economics and in the extension of the economic approach to other fields of social science, is however not an integral part of the *homo oeconomicus* concept. The central assumption of that concept, that each agent is concerned with the maximization of his own welfare, can be equally formalized in the utilitarian framework that has a long tradition in economics. The dominance that economics achieved in recent decades over the other social sciences involved the replacement of the *sociological conception of man* by the economic one. The essence of the sociological conception is that the individual identifies with the society in which he lives and is therefore directly concerned with the functioning and wellbeing of the society as a whole.

In an important book, the sociologist Lars Udehn (1996) has given a critical assessment of the economic approach to public choice, contrasting it with the sociological approach. He also discusses how and why the transformation occurred. I quote from his introductory chapter, omitting the extensive references cited in the text.

We have seen, over the last three decades, an immense spread of economic thinking in all branches of social science; a phenomenon commonly referred to as 'economic imperialism', or 'universal economics'. Outside its traditional domain, the economic approach has proved most useful in the analysis of politics. Since the 1950s, there has been an invasion of economists into political territory. Many political scientists, too, have become impressed by the apparent power of this perspective and adopted its tools. Being a discipline with a topic, but no particular approach, political science is easily subjected to the influence of neighbouring disciplines...

Before economics became the exemplary model, political science received a strong impetus from sociology and social psychology. There were three sources: behaviouralism, group theory and structural-functionalism. Of these, 'behaviouralism' was the most important and, in a wide sense of the term, included the other two. In a narrow sense, 'behaviouralism' signified an empiricist approach to political behaviour.

The movement known as 'behaviouralism' in political science – not to be confused with 'behaviourism' in psychology – was part of a larger movement in all social sciences. Its overriding aim was to turn the social sciences into hard sciences. This was to be achieved by systematic empirical study of human behaviour. Despite the transdisciplinary character of the movement, behaviouralism was to a considerable extent an importation from sociology, or its empiricist branch. The main instrument of research; the survey, came from sociology, as did some of the classic investigations...

At the end of the 1960s, however, political science went into a state of crisis, not as severe as that in sociology, but serious enough. Behaviouralism dissolved and was replaced by a post-behavioural theoretical pluralism. The strongest 'paradigm' eventually to emerge out of this plurality of voices was the economic approach to politics. Economics replaced sociology as the main external source of inspiration for political scientists. Great hopes were placed upon the new political economy – as twenty years earlier upon behaviouralism – as a means of making political theory more scientific or, at least, more explanatory. Behind these hopes lay an important development of the economic approach itself. In the 1940s and 1950s, it had become enriched by game theory, dealing with strategic behaviour – a phenomenon of some importance in political life.

Udehn is describing the effect of *scientism* on fields that are uncertain and defensive regarding their scientific status. By scientism I mean the belief that a field can become scientific by adopting some aspect of the methodologies of the natural sciences. The scientism connected with the adoption of the economic model focused on the use of a particular kind of abstract model, in conjunction with the use of mathematics, particularly game theory. The emphasis on the use of a particular method as an end in itself is characteristic of scientism. In contrast to this, genuine science begins with a relatively unstructured search for a *problem* that can be solved if a suitable method for dealing with it is found.

I turn to the movement in the opposite direction, from social science to economics. The subject is *happiness research*. Two excellent recent books by economists discuss this research and its implications for economics. They are Frey and Stutzer (2002) and Layard (2005). The field is interdisciplinary, originating in sociology and psychology and increasingly connected also with the medical – and neurosciences. The standard tool of this research is the survey questionnaire. The most important finding is that, beyond the basic necessities, material consumption is less important than the individual's integration in his community. This includes such elements as family, friendship, political participation and other aspects related to the quality of social life.

This research is sociological in all of its basic aspects: It is sociological in its basic findings that support the sociological conception of man. Its methodology, survey research is a standard sociological tool. Finally, the research is intensely empirical with very little by way of an analytical/mathematical theory structure.

A comparison of these two fields illustrates the fragmentation and ineffectualness of contemporary economics and social science. They deal with what is essentially the same subject, the representation and aggregation of individual preferences, yet, there is no communication between them. In happiness research the basic assumption is that subjective states can be measure and represented on a cardinal scale. In collective choice theory, the very possibility of doing this is denied. If on this issue the happiness researchers are right, as I am sure they are, then the entire field of collective choice is wasting its energy in vain attempts at solving a problem that has a trivial solution. That this is the case is the argument in Hillinger (2005). Furthermore, collective choice is the central, most massively researched problem in political science. If political scientists are unable to solve the problem of organizing a vote, than it stands to reason that they will be unable to offer guidance to any problem of political organization. This also seems to be the case.

The counterpart to the inability of an established discipline to recognize and incorporate the findings of a related new discipline is the inability of the new discipline to recognize the relevance of the problems for the solution of which the older discipline was created. Both Frey/Stutzer and Layard conclude their books with appeals to politicians to implement the utilitarian program by maximizing total utility as measured by the happiness researchers. This is very naïve and amounts to a denial of the need for either economic or political theory. It is hardly possible to increase average happiness without being able to predict the consequences of economic and social policies and this cannot be done without a valid economic theory. To give just one example: As this is being written, there is a debate among German politicians regarding the establishment of minimum wage legislation. The aim of this legislation, to improve the situation of the receivers of very low incomes, would certainly be approved by happiness researchers. Unfortunately, there is no agreement as to whether such legislation would have the intended effect. Critics argue that, by destroying jobs, the intended legislation would produce the opposite of the intended effect. Similar problems arise with almost any kind of legislation. Goodwill is thus not a sufficient condition for reaching the utilitarian optimum.

Similar problems arise at the level of political theory. The characteristic assumption of the public choice literature of the past decades is that politicians and other actors in the political

process are motivated, not by some notion of serving the general interest, but rather by the desire to maximize their own utility. The question of how to structure political institutions so as to serve, as well as possible, the general interest is a profound problem for which modern political theory provides little illumination. One might have expected happiness researchers to contribute towards a solution by insisting that voters should be allowed to express their preferences on a cardinal scale. But even such a modest attempt at transferring their most basic insight to the structuring of political institutions did not occur.

The fragmentation of economics and the social sciences into non-communicating segments has been increasing, driven by a number of factors. Most basic is the lack of agreed upon criteria for determining truth or relevance. When the lack of common criteria is coupled with the pressure to publish quickly in order to make a career, the consequence is that those entering the professions are under great pressure to quickly join some narrow group and accept their criteria in order to gain the acceptance of relevant journals and referees. This leads to a certain sectarianism and to a tendency of isolating the group from outside criticism. In economics, a further impetus towards fragmentation has resulted from the creation of bureaucratic institutions out side of the universities that were given, or assumed, the prime responsibility for major areas of empirical science. These are in particular the statistical agencies that have taken on the responsibility for economic statistics, not only at the level of actually producing the data, but also at the level of theoretical justification of the methods employed. The responsibility for macroeconomic policy has in large measure been assumed by central banks and international monetary institutions. The consequence is that these institutions strongly influence macroeconomic research. Bureaucratic governmental institutions develop their own cultures that are rather different from those prevailing in academia. Also, research at such institutions is likely to be dominated by their political agendas. For example, it is inconceivable that central bankers would seriously consider Friedman's proposal of letting the money supply grow at a constant rate, because this would drastically reduce their importance and prestige.

At present the economic and the sociological models of man are separated by two counterproductive divisions: a division between theory and empiricism and a division between the ideologies of the political Right and Left. These divisions make no sense since neither side can by itself explain all aspects of social systems. The consequence has been the oscillation in politics between Left and Right. Each ideology, when in power, fails to achieve its goals, opening the path for the ascent of the opposing ideology.

Both models have their origin in the work of Adam Smith; the economic model in *The Wealth of Nations* and the sociological model in *The Theory of Moral Sentiments*. Smith saw no conflict between the two; he applied each model to those aspects of society for which it is relevant. The embrace of the models by competing political ideologies and the extreme divide between theoretical and empirical work, while having their origins in his time, were later developments.

The ineffectiveness of current economics and social science is due to the fact that they have neither freed themselves from political ideologies, nor have they accomplished the close interaction between observation and theory that is characteristic of the natural sciences. The solution of the grave problems of the contemporary world desperately requires an objective and accomplished social science. Unfortunately, it is not in sight.

14. REFERENCES

- Arrow, Kenneth J., and Scitovsky, Tibor (eds.) (1969), *Readings in Welfare Economics*, Homewood, Ill., Richard D. Irwin.
- Balk, Bert M. (2000), Divisia price and quantity indices: 75 years after, Working Paper, Statistics Netherlands, Voorburg.

- Balk, Bert M. and Diewert, W. Erwin (2001), A characterization of the Törnqvist price index, *Economics Letters*, 72, 105-16.
- Bennet, T. L. (1920), The theory of measurement of changes in the cost-of-living, *Journal of the Royal Statistical Society*, 83, 455-462.
- Diewert, W. Erwin (1976a), Exact and superlative index numbers, *Journal of Econometrics*, 4, 7, 115-145. Reprinted in Diewert and Nakamura (1993).
- (1992), Exact and superlative welfare change indicators, *Economic Inquiry*, 30, 565-582.
- (1998), Index number theory using differences instead of ratios, Department of Economics, University of British Columbia, Vancouver, Discussion Paper 98-10.
- (2000), The quadratic approximation lemma and decompositions of superlative indexes, Discussion Paper 00-02, Department of Economics, University of British Columbia, Vancouver
- (2001), The consumer price index and index number theory: a survey, Discussion Paper No. 01-02, Department of Economics, University of British Columbia, Vancouver.
- Diewert, W. Erwin and Nakamura, Alice O. (eds.), (1993), *Essays in Index Number Theory*, Vol. 1, Amsterdam, North Holland.
- Dupuit, J. (1844), De la mesure de l'utilité des travaux publics, *Annales des Ponts et Chaussées*, 8. Translated and reprinted in: Arrow and Scitovsky (1969).
- Eatwell, John, Milgate, Murray, and Newman, Peter (eds.), (1987), *The New Palgrave: A Dictionary of Economics*, London, Macmillan Press.
- Eheman, Christian, Katz, Arnold J., Moulton, Brent R. (2002), The chain-additivity issue and the US national economic accounts, *Journal of Economic and Social Measurement*, 28, 37-49.
- Fisher, Franklin M. (1987), Aggregation problem. In Eatwell et al. (1987).
- Fisher, Irving (1927), *The Making of Index Numbers*, third ed. Boston, Houghton Mifflin.
- Harberger, Arnold (1971), Three postulates of applied welfare analysis, *Journal of Economic Literature*, 9, 785-797
- Hartley James E. (1977), *The Representative Agent in Macroeconomics*, London, Routledge.
- Hicks, John R. (1941-42), Consumers' surplus and index numbers, *Review of Economic Studies*, 9, 126-137.
- (1945-46), The generalized theory of consumers surplus, *Review of Economic Studies*, 13, 68-74.
- Hillinger, Claude (2001), Money metric, consumer surplus and welfare measurement, *German Economic Review*, 2, 2, 177-193.
- (2002), Consistent aggregation and chaining of price and quantity measures. *Journal of Economic and Social Measurement*, 28, 1-20.
- (2003a), The money metric, price and quantity aggregation and welfare measurement, *Contributions to Macroeconomics*, 3, 1, article 7, 1-34.
- (2003b), Output, income and welfare of nations: Concepts and measurement, *Journal of Economic and Social Measurement*, 28, 4, 219-237.
- (2005), The case for utilitarian voting, *Homo Oeconomicus*, 22, 3, 295-321. The paper can be downloaded at my SSRN homepage at:
http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=191500
- (2006), Science and ideology in economic, political and social thought. Downloadable at my SSRN home page at: <http://ssrn.com/abstract=945947>
- Hotelling, Harold (1938), The general welfare in relation to problems of taxation and of railway and utility rates, *Econometrica*, 6, 242-69. Reprinted in: Arrows and Scitovsky (1969).
- Jorgensen, Dale W. (1990), Aggregate consumer behavior and the measurement of social welfare, *Econometrica*, 58, 1007-1040.

- Kirman, Alan P. (1992), Whom or what does the representative agent represent? *Journal of Economic Perspectives*, 6, 2, 117-136.
- McKenzie, George W. (1983), *Measuring Economic Welfare: New Methods*, Cambridge University Press.
- Myrdal, Gunnar (1969), *Objectivity in Social Research*, New York, Pantheon.
- Petty, William (1691), *The Political Anatomy of Ireland*. Reprinted in: Hull, C. (ed.), *Economic Writings*, 2 vols. Reprinted, A. M. Kelley, New York, 1964.
- Pollak, Robert A. (1980), Group cost-of-living indexes, *American Economic Review*, 70, 2, 273-278.
- Quesnay, François (1759), *Tableau Economique*, Third edition, Paris.
- Robbins, Lionel (1932), *An Essay on the Nature and Significance of Economic Science*, London, Macmillan.
- Samuelson, Paul A., (1942), Constancy of the marginal utility of income. In: Lange et al. (eds.), *Studies in Mathematical Economics and Econometrics in Memory of Henry Schulz*, University of Chicago Press, 75-91.
- (1956), Social indifference curves, *Quarterly Journal of Economics*, 70, 1-22.
- Samuelson, Paul A. and Nordhaus, William D. (1985), *Economics*, 12th Edition, New York, McGraw-Hill.
- Slesnick, Daniel T. (1998), Empirical approaches to the measurement of welfare, *Journal of Economic Literature*, 36, 4, 2108-2165.
- Spiegel, Henry W. (1971), *The Growth of Economic Thought*, Englewood Cliffs, N. J., Prentice-Hall.
- Stone, Richard (1978), Keynes, political arithmetic and econometrics, *Proceedings of the British Academy*, 64, 55-92.
- Takayama, Akira (1987), consumer surplus. In Eatwell et al. (1987).
- Theil, Henri (1967), *Economics and Information theory*, Amsterdam, North Holland. (1968), On the geometry and the numerical approximation of cost of living and real income indices, *De Economist*, 116, 6, 677-688.