Why Friedman's methodology did not generate consensus among economists?

David Teira Serrano

Dpto. de Lógica, Historia y Filosofía de la ciencia (UNED) and Urrutia Elejalde Foundation

ADDRESS:

Dpto. de Lógica, Historia y Filosofía de la ciencia
UNED. Edificio de Humanidades
Paseo de Senda del Rey 7
28040 Madrid (Spain)
Phone: (34) 913988392
Fax: (34) 913987693

E-mail: dteira@fsof.uned.es

ABSTRACT:

In this paper I study how the theoretical categories of consumption theory were used by Milton Friedman in order to classify empirical data and obtain predictions. Friedman advocated a case by case definition of these categories that traded theoretical coherence for empirical content. I contend that this methodological strategy puts a clear incentive to contest any prediction contrary to our interest: it can always be argued that these predictions rest on a wrong classification of data. My conjecture is that this methodological strategy can contribute to explain why Friedman's predictions never generated the consensus he expected among his peers.

.

Why Friedman's methodology did not generate consensus among economists?

1. INTRODUCTION

That scientists behave as disinterested truth seekers was a common assumption among many classical sociologists of science (after Merton) and remains a basic, but often implicit, tenet of many contemporary philosophers. According to these assumption, scientists would put their personal interests aside during the empirical assessment of any hypothesis. For instance, many theories of evidence assume that by accumulating evidence, even the most self-interested scientist will have to admit, sooner or later, that either the hypothesis is confirmed or disconfirmed independently of her own preferences. Yet there is no precise methodological mechanism to account for this decision: it is a matter of trust in the personal integrity of every scientist (e.g., Nagel 1961, pp. 494-95).

However, there are many disciplines, especially among the social sciences, in which the accumulation of evidence has not yielded a broad theoretical consensus. It is often argued that personal biases of various kinds prevent many social scientists from attaining consensus, but the underlying rationale (their personal incapability to overcome such prejudices) is no more convincing than the aforementioned appeal to integrity. Indeed, when the incentives are powerful enough, scientists seem to act in their own interest irrespectively of their field (e.g., Krimsky 2003). In this paper I will explore through a case study an alternative explanation as to why some economists disagreed more than their peers in the natural sciences, in spite of the evidence accumulated.

The intuition that I will develop is this. In the natural sciences, theoretical concepts are considered as unified semantic entities which can be systematically applied in a

diversity of contexts[1]. We cannot manipulate these concepts at will in order to avoid evidence contrary to our interests, and hence exclude these items of contrary evidence from the definition. We may rather expect these categories to filter such partial interests through an iterative procedure of self-correction, as Hasok Chang (2004) recently defended in the case of temperature. Thermologists assumed that, independently of the measurement procedure, they were all measuring different manifestations of the same phenomenon, whose observability excluded optional interpretations. Therefore, they were compelled on methodologically grounds to reconcile their results and make them converge under the same theoretical concept (*temperature*). Even if they had to rely on various conventions, which were arbitrary to a certain extent, the accumulation of thermometrical evidence compelled thermologists to render all their measuring scales *coherent*, independently of their personal preferences. Coherence acted thus as a built-in methodological mechanism that screened biases and yielded consensus.

It seems as if such coherence cannot be so easily attained in the social sciences. At least, this has been a traditional dilemma of the most mathematised (and, in that respect, more coherent) of them, neoclassical economic theory. The more definitional constraints we impose on its categories, the more difficult it is to capture data and increase its empirical content. Therefore the accusation of *formalism* against neoclassicism. Yet, if we relax the demand for coherence of economic categories in order to allow for their wider application, disagreement as to their proper use seems both inescapable and difficult to overcome. Natural and social scientists may be equally prone to bias (i.e., selecting evidence in according to their particular interests), but there might be no built-in mechanism in the latter's categories to prevent it.

In this paper, I will illustrate this situation with a case study of an outstanding economist who also became famous for a methodological defence of the role of evidence in the attainment of a scientific consensus in economics. According to Milton Friedman (1953a), economic theory could be considered a scientific discipline inasmuch as it provided a "language" and "a body of substantive hypothesis" that jointly yielded more accurate predictions, independently of the realism of the assumptions employed to

---

[1] See Hempel 1952 for a classical formulation. A recent illustration of this could be the *constraints* which are postulated in the structuralist reconstruction of scientific theories to guarantee the unity of the various applications of a given concept between different models. See Balzer et al. 1987.

obtain these predictions. Friedman's methodology was originally intended as an exposition of the principles that guided his own research (Hammond 1998) and was subsequently adopted by many others in the profession. Paradoxically, in the years to come Friedman turned out to be a controversial economist, whose predictions did not produce much the desired consensus among his peers in spite of their purported accuracy.

Instead of claiming that ideological biases can blind us to acknowledge the success of our adversaries' predictions, as Friedman himself almost came to suspect (Friedman & Friedman 1998, p. 219), I will argue that Friedman's methodological prescriptions concerning the definition of the categories of consumption theory relaxed coherence and hindered theoretical consensus. I will discuss first (§2) how Friedman advocated a definition of the categories of consumption theory on a case by case basis in accordance with the interests of the analyst. His underlying motivation was to allow for as many data classifications as possible since this was required to apply Friedman's statistical technique of choice (linear regression) and obtain predictions (without which the scientific status of economics would seem questionable). There was thus a clear tradeoff between predictions and coherence: any attempt to encompass all the possible applications of a single category under a well articulated definition (as in general equilibrium theory) prevented the accurate description of each particular market analysed and the acquisition of significant predictions about them[2].

Yet Friedman's definitions became easily contestable by anyone who disagreed with either the assumptions or the consequences of the analysis, no matter how accurate the predictions. It was always possible to argue that a different classification of the data was equally cogent. In other words, consumption categories interpreted *à la* Friedman, were not coherent enough to filter personal biases and yield scientific consensus. I will analyze a couple of studies that Friedman considered methodologically exemplary in

---

[2] Philippe Mongin (1988, p. 311) already noticed how threatening this methodological strategy was. Even if modifying definitions was sometimes called for in a particular aplication in view of achieving a particular goal, he argued that each modification should be later incorporated into the logical syntax of the theory —e.g., by means of auxiliary hypotheses— if its conceptual unity is to be preserved in the testing process.

order to see how the objections concerning the inability of his results to generate consensus were articulated. First (§3), we will discuss some chapters from his first book, *Income from Independent Professional Practices*, coauthored with Simon Kuznets, and then (§4) the central hypothesis of *A Theory of the Consumption Function*. Friedman's approach will appear here as a less inspiring methodological alternative than it is sometimes deemed to be. To conclude, I will discuss to what extent we can attribute the lack of scientific consensus (and therefore theoretical progress) to a lack of definitional coherence.

2. CLASSIFICATIONS AND PREDICTION

The meaning of scientific terms, as used by a given community, is either determined by sense data or by the shared interests of its members. Whereas in the first approach the meaning of a term is fixed (in various degrees) by its reference, in the second one it is rather agreed (more or less explicitly) between its users. In both cases, it is assumed that such determination is robust enough as to secure a certain coherence through all the applications of each term. In this respect, no individual can change at will the meaning of a term in accordance with her particular interests, unless these correspond to either the very nature of its reference or the interests of the rest of the community. Both alternatives have been widely explored by philosophers and sociologists alike.

Less attention has been paid to those scientific terms in which the data do not allow for a uniform categorization, or the interests of the community do not converge in a single definition. This apparently was the case in demand studies when Friedman produced his methodological manifesto. There were conflicting interpretations of demand theory, namely hinging on the general vs. partial equilibrium divide, the former being less prone than the latter to yield empirical predictions. Friedman's methodology for positive economics dictated that we should choose among these interpretations in accordance with the accuracy of the predictions they yielded.

In Post War America, a generation of economists educated in the aftermath of the Great Depression were eager to contribute to a well-ordered economic governance, and it was quite a widespread opinion that good predictions were the best means to facilitate such improvements (e.g., Despres et al. 1950). There was an echo of this concern in Friedman's methodological essay: if the American economic profession could reach an

agreement on which theories were better, their fellow citizens would be able to make an informed decision regarding the different economic policies available. They would be able to choose the one whose expected output (as predicted by positive economics) better suited their interests.

However, in order to obtain these predictions, agreement on the proper theoretical categories to be used was crucial. These categories should allow us to classify data and this was crucial to the application of Friedman's chosen statistical techniques. Regression analysis, as Ronald Fisher taught it, relied explicitly on a classificatory device: the contingency tables invented by Karl Pearson (Porter 1986, pp. 311-14). These tables collected observations in their rows and columns and delivered, in the bivariate case, an intuitive test of their fitness to the distribution pattern required by correlation analysis so that further predictions could be derived through a regression. Yet the classification of data in these tables supposed agreement concerning the definition of the featured categories, so that we could place data in the table without any ambiguity.

Friedman's statistical training was conducted first under the supervision of a disciple of Pearson (the econometrician Henry Schultz) and, most significantly, at Harold Hotelling's department at Columbia, one of the very few places in the United States where the theoretical developments refinements introduced by Ronald Fisher were taught and applied to economic issues. The statistical methods that Friedman applied throughout his career were in many ways Fisherian and he was a firm believer in the virtues of simple regressions (Teira 2007). We may safely assume that the predictions he had in mind when drafting his methodological piece were yielded by these methods. What type of theoretical categories were more suitable for this approach?

In consumption and income studies, the classification of data depended on categories which had been theoretically defined during the neoclassical revolution in the last quarter of the 19th century –though various versions coexisted, depending on times and places, during the following fifty years. While he was still an undergraduate, first in Rutgers and later in Chicago, Friedman was exposed to different readings of Leon

Walras and Alfred Marshall[3]. According to the interpretation of the former advocated by Henry Schultz, the equations that articulated general equilibrium theory provided a complete picture of the causal interrelations that articulated the economic system as a whole. Through regression analysis, claimed Schultz, all the relevant links between such factors could be empirically traced to test the theory (Teira 2006). Friedman took sides with his economic mentor, Arthur Burns, who granted with Marshall that such an exhaustive representation was admissible in principle, but empirical demand analysis required the division of the causal nexus to focus only on the relevant variables in the market under consideration (Hammond 1998). Only after a set of partial solutions had been obtained would the Marshallian theorist consider its combination "into a more or less complete solution of the whole riddle". In Friedman's own words, Walras would have provided an "idealized picture of the economic system", but not an "engine for analyzing concrete problems", as did Marshall (Friedman 1953b, p. 57).

Moreover, Friedman argued that the Walrasian stance was *sterile* for predictive purposes. Defining each theoretical category in a consistent way throughout the entire economic system required a logical rigor that hindered its empirical application. The classification of market data required analytical *filing cases à la* Marshall:

> Demand and supply are to him [Marshall] concepts for organizing materials, labels in an "analytical filing box." The "commodity" for which a demand curve is drawn is another label, not a word for a physical or technical entity to be defined once and for all independently of the problem at hand. (Friedman 1953b, p. 57)

The categories of demand theory were thus classificatory devices to organize empirical data, without the constrictions imposed by the, so to speak, *causal rigidness* of the Walrasian definitions. To put this in our own terms, the reference of these categories could not be theoretically determined, since the causal connections were far too intricate to be of practical use in the classification of data. Without such classification, as we know, there would be no regressions and hence no predictions. Demand theory would

---

[3] Notice that I do not intend to judge whether Friedman or any other author herein discussed was really faithful to the teachings of either Walras or Marshall, but take at face value their own claims. My position in this respect is closer to Hammond 1996, pp. 26-45

not reach thus the status of positive economics. Therefore, a clear methodological rationale existed for the treatment of the categories of demand theory as analytical filing boxes.

To illustrate the divergence between the Marshallian and Walrasian approaches, let us consider the following example extracted from a 1943 joint paper by Friedman and Allen Wallis on the empirical derivation of indifference functions (Wallis & Friedman 1942). We depart from a given set of statistical data and we want to analyse the quantitative relation of consumer expenditures to price and income, in order to predict the effect of any change in these on consumption. The definition of an indifference function obliges us to classify the data "once and for all", by using concepts such as goods, taste factors and opportunity factors. Yet, argued Friedman, the same items may be arranged under any of these three concepts depending on the particular circumstances "of the problem in hand". This is why "it is so difficult to specify reasonable conditions for deriving indifference curves from observational data" (Wallis & Friedman 1942, p. 187). And even in those cases when the obstacle was overcome, the Walrasian approach was usually incapable of yielding correct predictions on data other than those used in their empirical specification, since they lacked "elasticity" to incorporate other data. Rather than wrenching the observable data into a pre-existing theoretical scheme, Friedman advocated a reformulation of our theoretical categories to generalize them (Stigler 1994, p. 1200).[4]

Now, we must also notice that Friedman's interest in predictions was not strictly epistemological. A pure mathematical analysis of the interrelations between the different sections of an economy, such as Walras', would be of theoretical interest per se. But if we wanted economics to play a role in the improvement of social welfare, as many economists of Friedman's generation did, it was not necessary (they argued) to grasp the complete causal structure of those interdependencies in order to decide how

---

[4] Though he remained the archetypical positivist throughout the second half of the 20th century, we must notice how far he was from the Vienna Circle as to the proper definition of theoretical terms. Whereas in the latter their logical articulation "once and for all" was the only way to assess its empirical content, Friedman would argue against axiomatization to preserve the predictive fruitfulness of economics. It is not strange that the Walrasian approach deserved more attention on the part of the Viennese tradition.

to act (Despres et al. 1950, p. 512). If accurate predictions concerning the consequences of our policies were available, we could attain some rational consensus on which one to choose: "the role of statistics is to resolve disagreement among people. It's to bring people together" (Friedman quoted in Hammond 1993, p. 225)

We now have all the ingredients in our analysis: a methodology that relaxed coherence in the definition of the categories of demand theory for the sake of predictions and scientific consensus, which were desired for a mixture of epistemological and practical concerns (though we may well conjecture that the most compelling incentive was provided by the latter). Here is the dilemma: Schultz's Walrasian framework provided, once and for all, the kind of tight theoretical structure needed for making people converge on their predictions, but these latter were difficult to obtain. Friedman's Marshallian framework required a previous agreement on the part of the economic profession concerning the definition of the categories of demand for each particular problem. But if such definitional agreement was not reached, the Marshallian economists would find themselves in a difficult position: there would always be a rationale to contest the underlying classification as based on a wrong definition of the theoretical categories involved. In the next two sections we will see how these classifications could work unsanctioned, yet at the cost of not yielding the consensus that Friedman strived to attain.

3 INCOME FROM INDEPENDENT PROFESSIONAL PRACTICE

> So far as my own work is concerned, I should not want to judge its importance, but I do feel that *Income from Independent Professional Practice* (written jointly with Simon Kuznets), particularly chapters 3 and 4, embodies the appropriate methodological approach in respect to the combination of empirical and theoretical analysis." (Friedman to E.B. Wilson, 1946 *apud* STIGLER 1994, p.1200).

Income from independent professional practices was the result of Friedman's first assignment at the NBER, which he joined in 1937 in order to work on income distribution in the United States with Simon Kuznets (Friedman & Friedman 1998, p. 68). Having completed a first draft, Kuznets asked Friedman to revise and extend it. The final version was finished about 1941, though it would take four more years to get it published. In 1946, Friedman was awarded his doctoral degree at Columbia. His dissertation consisted of a manuscript derived from the book. Let us focus on the two

chapters that Friedman found methodologically most satisfactory and a clear anticipation of his 1953 essay.

Friedman is responsible for the third chapter and the second section of the fourth, where we found the analyses that I will here discuss (Friedman & Kuznets 1945, p. xii). These two chapters consist of 110 pages organised as follows: a thorough statistical discussion of income data drawn from five different professions is carried out in the first 90 pages, while the remaining 20 focus on a demand theoretical analysis of the differences in average income between two of the professions, medicine and dentistry. The first part delivers a sort of analytical description of the data from which a hypothesis is derived: professional workers "constitute a "noncompeting" group" as their number (and hence their income) is not exclusively determined by the "relative attractiveness of professional and nonprofessional work", but by the number of prospective students who count on the particular resources required to pursue the career in question (Friedman & Kuznets 1945, p. 93). The second part addresses a more specific issue, whether *licenses* may restrict competitive access to medical practice and therefore account for the observed differences in average income between physicians and dentists. Given that both professions require similar abilities and training, we should expect the prospective practitioners to choose between them mainly on the basis of their respective "level of return" (Friedman & Kuznets 1945, p. 123). The theoretical issue at stake is whether their respective average income corresponds to the *equilibrium level*:

> [I]f entry into the two professions were equally easy or difficult, one might expect an adjustment of the levels of return in them that would equalize their net attractiveness in the eyes of a considerable fraction of those in a position to choose between them (Friedman & Kuznets 1945, p. 124)

The way the issue was addressed exemplifies what I take to be the methodological core of Friedman's Marshallian approach. In order to render the initial statistical analysis significant in a demand theoretical setting, and so to decide whether the markets for dentists and physicians were in equilibrium, the theory was to be reformulated so as to "generalize the observable data". This was accomplished in two methodologically crucial steps.

First, Friedman took the *arithmetical mean* measuring the level of income as a proxy for the utilitarian considerations that theoretically account for the choice of profession

by prospective practitioners. The expected mean income would determine individual decisions concerning which career to pursue in order to balance its expected costs (Friedman & Kuznets 1945, pp. 65, 145). Choosing according to this "summary figure" would yield a theoretically significant result: even if the individual is unaware of such a statistical decision rule (Friedman & Kuznets 1945, p. 96), prospective entrants as a whole would decide their career in such a way as to attain a negatively sloped demand curve in accordance with utility theory[5]. No rationale is provided to justify such a step, though Marshall's authority may be certainly be invoked at this point (Friedman 1949).

Friedman's second step consisted in stating a statistical definition of the quantities and prices of the commodities that dentists and physicians sell, since the relevant "unit of service" as well as its cost —so easy to state theoretically— were in fact far from obvious (Friedman & Kuznets 1945, p. 155). As for prices, they could only be approximated by the global amount of money the consumers as a whole are willing to spend for those services (Friedman & Kuznets 1945, p. 158). Since this partially depends on the number of practitioners supplying them, a demand curve for medical services will be traced in which the "price" will correspond to the average net income per doctor and the "quantity" to the number of them operating in a given market once the appropriate *ceteris paribus* clauses are arranged (*ibid.*). Correspondingly, the supply curve of entrants will be drawn by taking the practicing professional as the "unit of service" and his expected mean income (*ceteris paribus*) as its "price" (Friedman & Kuznets 1945, p. 156). However, the data did not allow for a curve to be drawn (p. 161).

Therefore, to deal with theoretical categories as *filing cases* amounts to *equating them with certain available statistical measures* which classify the data in a way that is considered relevant "in view of the problem at hand". No theoretical constraint on the definition of these categories is to be taken into account, even though the resulting

---

[5] "But if few or no individuals go through the reasoning or calculation underlying our estimate, many do try to take account in some way of the differential costs attached to the choice of one profession rather than another. Implicitly or explicitly, they do attempt to estimate the differences in incomes that will compensate for this costs. It seems reasonable to suppose that they are as likely to overestimate as to underestimate; and, on the whole, we may expect the estimates to cluster about the correct value." (Friedman & Kuznets 1945, p. 127)

concept may seem dubious from a purely analytical standpoint. However, it might be justified on the basis of the practical interest of the results so obtained.

> The analysis is necessarily conjectural and our quantitative results are only a rough approximation. But the problem is real, and a rough approximation seems better than none (Friedman & Friedman 1998, p. 8).

To see how rough this approximation was (and to what extent it justified Friedman's methodological strategy), take for instance the discussion of the income equilibrium level for dentists and physicians, which was conducted by Friedman and Kuznets by means of regression analyses performed separately on the 1934 and 1936 data for the following variables: per capita income of both professions, their number per 10.000 population and general per capita income (Friedman & Kuznets 1945, pp. 161-173). The equilibrium difference in income between the two would be the one that made both professions equally attractive to prospective entrants, taking into account the relative demand for their respective services. However, the regressions were not good enough to discern what particular rise in the ratio of physicians to dentists would be needed to reduce the ratios of their incomes to a level close to equilibrium (which was informally estimated to be about a 17 per cent difference: Friedman & Kuznets 1945, pp. 132, 172). Given that the observed difference amounted to a 32 per cent, the authors could wonder whether it was to be explained as a result of "free and moderately rational choice", or of an entry barrier in the form of medical licensing instead (Friedman & Kuznets 1945, pp. 125).

Even if such statistical analysis might seem in many ways inconclusive, Friedman and Kuznets suggested that the *American Medical Association* may be guilty of monopolistic practices. This sparked a controversy within the NBER delaying publication of the book for several years (Friedman & Friedman 1998, p. 74). It finally went to print with a "Director's comment" by Carl Noyes (Friedman & Kuznets 1945, pp. 405-410), whoe stated certain "reservations" over the scientific validity of the results obtained in chapters 3 and 4. Noyes argued that these results could have been different had the authors' introduced additional distinctions (e.g., specialists vs. general practitioners) in their analysis. In other words, if they had used a *different classification of the data*. Indeed Noyes' concern was shared by some reviewers who nevertheless

praised the statistical effort undertook by the Kuznets and Friedman (in the spirit that a prediction is always better than none):

> This reviewer tends to agree with a comment made by C. R. Noyes to the effect that the assumptions made about the differences between these two professions were of such an arbitrary nature as to render highly untenable the statement that doctors restrict entry into their profession more than dentists. (Anderson 1946, 400; see also Barna 1947)

To put all this in our terms, the reference of the categories of demand theory in this market was either too general (the "unit of service" of the commodity) or simply unobservable (utility), and hence could not be discerned directly in the data. Friedman and Kuznets opted for identifying various statistical indexes which provided a data classification intuitively plausible and good enough for predictive purposes. We cannot discern whether any sort of personal bias (for or against the *American Medical Association*) motivated either our two authors or the NBER director but, if it existed, the predictions so obtained could not filter those biases out. Other classifications according to the interests of the analyst were no less legitimate and, in case of disagreement, demand theory could not dictate which one provided a more coherent interpretation.

However, *Income from independent professional practices* did not set a research agenda whose further development could have provided some sort of test of our conclusion. This will be now found in our second text under analysis, *A theory of the Consumption Function*.

4. A THEORY OF THE CONSUMPTION FUNCTION

> *A Theory of the Consumption Function*, published in 1957, comes closer than anything else that I have written to adhering faithfully to the precepts of my essay on methodology. That is one, but by no means the only, reason why I have long regarded it as my best purely scientific contribution, though not the most influential. (Friedman & Friedman 1998, p.222)

In 1951 Friedman wrote a four-page manuscript in which he developed a hypothesis on consumer behavior which would allow for the analysis of aggregate data on consumption (Friedman 1957, p. 13). From this seed would spring, six years later, *A*

Theory of the Consumption Function. In this section, we intend to develop a methodological analysis of the formulation of this *theory*, with a particular focus on its alleged methodological superiority with respect to *Income from Independent Professional Practices*.

The central element in the hypothesis is a result of Friedman's studies under Kuznets: the distinction between *transitory* and *permanent* aspects in the income data (Friedman & Friedman 1998, p.225). Rather than conferring a statistical interpretation on concepts already available, Friedman would coin new ideas on the basis, again, of a pre-existing statistical distinctions.

Let us firstly introduce the theoretical concepts that Friedman took as his starting point. The consumption function establishes the relationship between aggregate consumption, or aggregate savings, and aggregate income (Friedman 1957, p.17). Formally, this is a function $c_1 = f(W_1, i)$, where the variables stand for the *wealth in a given year* (W_1) and the *interest rate* (i), *wealth* being in turn a function of its *income* —the *expected income in a year*— and again the interest rate.

From a theoretical standpoint, it is supposed that the agent's preference can be represented along the distribution of his consumption by means of indifference curves. To the usual premises (negative slope, convexity) Friedman adds the *absence of a temporal preference* (Friedman 1957, p.28) in order to obtain a particular form of the function for *permanent consumption* c_p:

$$c_p = k(i, w, u) \cdot iW$$

Where *u* stands for the "utility factors", *w* for the proportion between *nonhuman wealth* and *permanent income* (Friedman 1957, p.17). The intuition underlying this additional assumption is that individuals do not usually consume according to the short-term fluctuations in their income, but rather adjust it to a certain average, which would be *the permanent income*. This theoretical concept cannot be directly equated to the current receipts which measured income in most statistical studies. (Friedman 1957, p. 10). Moreover, the income earned by a consumption unit during a certain period of time is by definition *unobservable*. According to Friedman, if we consider permanent income to be the quantity which a consumption unit might expend without its wealth being thereby affected (Friedman 1957, pp. 17, 20-21), we must admit – even though it may seem *embarrassingly obvious* (Friedman & Friedman 1998, p.225)– that such quantity

will depend on the income which the unit expects to earn in a period of time somewhat longer than that regarded by the data. Now this is an *expectancy* which is not directly reflected in these data[6]. And the same applies to permanent consumption. Two of the variables which articulate the hypothesis on the relationship between consumption and income are therefore *unobservable for any individual consumer unit* (Friedman 1957, p. 20)

So far, we may analyze Friedman's hypothesis as a purely theoretical result, akin to the Hicksian concept of income (among others). But we may also appraise it as a Marshallian reformulation of an existing theory to generalize certain patterns observed in the data. Friedman's distinction between permanent and transitory components in income was inspired, as he explicitly acknowledged, by a review published by his statistical mentor, Harold Hotelling, in 1933 (Friedman & Kuznets 1945, p. 331n; Friedman 1957, p.53n; see also Stigler 1996). Hotelling diagnosed a regression fallacy committed by Horace Secrist in *The Triumph of Mediocrity in Business*: to affirm the existence of a convergent trend between the respective means of several sets of data on annual business profit. These sets had been arranged in accordance with the initial profit of each firm, so that their respective means in successive years will be conditionalised on this value and will tend to the overall mean of the data. In other words, the variance among means of groups will diminish, since the transient component (exceptional good or bad profits) will be cancelled throughout the years coming closer to (*regressing towards*) the overall mean. Therein lies the fallacy of affirming this purely statistical convergence as an economic fact.

Fisher's analysis of variance provided a technique to decompose intraclass and interclass correlation, which also provided a significance test for linear regressions. Friedman learned about the technique from Hotelling in Columbia, or directly from Fisher's *Statistical Methods for Research Workers* (which Friedman considered "the serious man's introduction to statistics"). If we had to deal with a series of correlated measurements of a given variable fluctuating stochastically over time, classified into different arrays, the analysis of variance yielded a statistically relevant tool to improve the predictions that we may thereby obtain.

---

[6] "The adjective "planned" would perhaps be more appropriate in the present context than "permanent"." (FRIEDMAN 1957, p.11)

Friedman may have well wondered if there was a theoretical counterpart for this statistical classification in economic theory. Income was certainly one such variable, and it made perfect sense to decompose it into a permanent and transient part to analyze its fluctuations. Yet no assumption was made as to the particular mechanism that would have allowed each particular individual to carry out such decomposition. What mattered most was to read the aggregate data series "*as* if they had regarded their income and their consumption as the sum of two such components, and *as if* the relation between the permanent components is the one suggested by our theoretical analysis" (Friedman 1957, p.41). Conforming to his methodological stance, Friedman viewed the predictions that we could obtain as the only justification needed to grant such assumption, independently of both its realism and of the particular theoretical connections that allowed for the derivation of the hypothesis.

But these predictions, again, depended on the classification of the data that we could obtain on the basis of this decomposition. And, as Friedman occasionally acknowledged, the division of permanent and transitory income "is, of course, in part arbitrary, and just where to draw the line may well depend on the particular application". Marshallianism was here once more affirmed. Indeed, the central assumption as to the empirical testability of the hypothesis is that the correlation between the transitory components of both income and consumption is null (Friedman 1957, p.27; Hirsch & De Marchi 1990, p. 201), which implied that that transitory income does not affect the consumer unit's planned consumption. But what counts as permanent or transitory? As Chao observes,

> Friedman argued that in fact consumers are able to use their windfall receipts on planned consumption, whereas windfall receipts are usually statistically recorded as transitory income. Friedman concluded that unexpected windfall receipts should be regarded as transitory income and expected receipts regarded as permanent income (Chao 2003, p. 87).

The problem is that to identify such distinctions in the data in a non-arbitrary way we should be rely on a theoretical model of how individuals form expectations, allowing for some sort of precise empirical control. But this is precisely the sort of theoretical flexibility that "as if" reasoning grants and the Marshallian stance justifies. Friedman's methodology allows us to identify theoretical categories and statistical indexes without theoretical constrictions, transforming thus the former into filing boxes according to our

particular needs. Yet such needs may vary. Hence the predictions themselves will not reconcile our differences if we are led to question our classificatory assumptions.

Indeed some reviewers initially noticed that the intuition captured in the hypothesis was very plausible and the ensuing statistical analysis deserved all praise, but the theoretical underpinnings of Friedman's model were much more dubious (e.g., Champernowne 1958, Schultze 1958). Nonetheless, it was unanimously considered as worth developing, which subsequently happened in the following decades. We will not follow in detail the debate on the empirical tests the model went through, but will rely instead of the overall assessment provided by Tom Mayer almost twenty years later. As he put it (Mayer 1972, p. 6), the testing of the different theories of the consumption function had not yielded any significant consensus. Most economists disregarded those tests previously published by others "in the quaint belief that one's own test —being one's own— obviously deserves more credence". Mayer tried to survey and ponder all the evidence gathered by both supporters and critics in order to ascertain which theory was better. He concluded that the main implications of Friedman's hypothesis were not satisfactorily proven, the reason being:

> [S]omeone who insists on really rigorous confirmation or disconfirmation must accept the agnostic verdict that for now this hypothesis is neither confirmed nor disconfirmed. But if one is willing to work with the conventional definition of consumption, and to accept the assumption that the yield on transitory income is less than, say 50 percent -hardly a wild assumption- then there is evidence [...] that the transitory income elasticity of consumption *is* greater than zero (Mayer 1972, p. 348)

"The conventional definition of consumption" was based on the purchase of consumer goods as reported in the national income accounts, whereas others who advocated a theoretically grounded definition referred only to the actual destruction of consumer goods by use, not to the reported purchase of consumer goods (Mayer 1972, pp. 12-15). The problem was that no relevant data were then available to test this, and it might be argued that the durable goods included the consumer expenditures accounted for the positive correlation with transitory income. Herein lies the importance of Mayer's initial caveat.

We are again caught in a methodological trap: the accumulation of empirically testable predictions will not produce a consensus if there is no previous agreement on the

classification of data. But if we opted for more rigorous definitions to attain such classificatory agreement it might happen that we would be unable to classify the data from which the predictions could be derived. In the empirical analysis of Friedman's hypothesis it was customary to use both definitions of consumption, one in the formal exposition of the theories and the other in much of the testing of them. Therefore, it was always optional for the reader, according to her particular interests, to appraise the tests as if they were probing "a series of propositions which differ from these theories [e.g., permanent income theory (DTS)] in their definition of consumption".

5. CONCLUDING REMARKS

Friedman's methodology did not guarantee coherence in the definition and application of the categories of demand theory, but rather praised their disunity. I contend that this is why it failed to yield a wide professional consensus. My conjecture is that all sorts of interests might have prevented this: a long as there was an incentive to disagree, Friedman´s methodological stance delivered a Marshallian argument to justify it. Loose definitions are acceptable as long as they allow us to grasp the specificity of market data and obtain predictions therefrom.

The debate on Friedman's methodological essay has focused mostly on matters of epistemic principle: namely, whether his case against the realism of assumptions was philosophically sound, either as it stood or combined with other philosophical ideas (e.g., falsationism, pragmatism, etc). Friedman's methodological ambiguity is proverbial, and his self-proclaimed Marshallianism is no exception in this respect. As Hirsch and De Marchi (1990, p. 37) warn, we may not be able to derive a consistent interpretation of the Marshallian-Walrasian distinction if we take it as it appears in Friedman's works. Yet if we connect it to the statistical techniques he used to obtain predictions, the idea of treating theoretical categories as *filing cases* seems unambiguous: depending of whether our definition of these was particular or general, we would be able to obtain different classifications of the data.

Why economists should agree on a particular classification was one of the many questions that Friedman left unaddressed. He may have thought that most economists were as disinterested as he was (Friedman & D. Friedman 1998, pp. 218-19), but he was providing methodological arguments for precisely those who were not so they could justify their disagreement. In this respect, Friedman was wrong if he thought that the methodology he was advocating was similar to that of physics (1953a. pp. 16-19). If we

take Chang's analysis (2004) as counterexample, we see that the coherence of the theoretical categories in use is crucial to the progress of empirical measurement even when we do not have direct access to the world. Whether the Walrasian approach could do it better is a methodological question that we should leave open here.

6. REFERENCES

Anderson, R. L. "Review of Income from *Independent Professional Practice*." *Journal of the American Statistical Association* 41 (1946): 398-401.

Balzer, W., U. Moulines, and J. D. Sneed. *An Architectonic for Science*. Dordrecht: Reidel, 1987.

Barna, T. "Review of Income from *Independent Professional Practice*." *Economica* 14 (1947): 66-68.

Champernowne, D. G. "Review of *a Theory of the Consumption Function*." *Journal of the Royal Statistical Society* 121 (1958): 124-26.

Chang, Hasok. *Inventing Temperature. Measurement and Scientific Progress,*. N. York: Oxford University Press, 2004.

Chao, Hsiang-Ke. "Milton Friedman and the Emergence of the Permanent Income Hypothesis." *History of Political Economy* 35 (2003): 77-104.

Despres, E., M. Friedman, A. Hart, P. Samuelson, and D. Wallace. "The Problem of Economic Instability." *American Economic Review* 40 (1950): 505-38.

Friedman, M. "The Marshallian Demand Curve. " *Journal of Political Economy* 57 (1949): 463-95.

———. "The Methodology of Positive Economics." In *Essays in Positive Economics*, edited by Idem, 3-43. Chicago: Chicago University Press, 1953a.

———. "The Marshallian Demand Curve." In *Essays in Positive Economics*, edited by Idem, 47-99. Chicago: Chicago University Press, 1953b.

———. *A Theory of the Consumption Function*. Princeton: NBER, 1957.

Friedman, M., and R. D. Friedman. *Two Lucky People. Memoirs*. Chicago: Chicago University Press, 1998.

Friedman, M., and S. Kuznets. *Income from Independent Professional Practice*. N. York: NBER, 1953.

Hammond, J. D. "An Interview with Milton Friedman on Methodology." In *The Philosophy and Methodology of Economics. Vol. 1*, edited by B.C. Caldwell, 216-38. Aldershot: Edward Elgar, 1993.

Hammond, J. Daniel. "Friedman, Milton." In *The Handbook of Economic Methodology*, edited by J. B. David, D. W. Hands and U. M鉵. Cheltenham: Edward Elgar, 1998.

Hempel, C. *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press, 1952.

Hirsch, Abraham, and Neil De Marchi. *Milton Friedman: Economics in Theory and Practice, The Economists*. New York: Harvester Wheatsheaf, 1990.

Hotelling, H. "Rese駄De H. Secrist, *the Triumph of Mediocrity in Business*." *Journal of the American Statistical Association* 28 (1933): 463-65.

Krimsky, Sheldon. *Science in the Private Interest*. Lanham: Rowman & Littlefield, 2003.

Mayer, T. *Permanent Income, Wealth and Consumption*. Berkeley: University of California Press, 1972.

Mongin, P. "Le R閙isme Des Hypoth鑣es Et La Partial Interpretation View." *Philosophy of the Social Sciences* 18 (1988): 281-325.

Nagel, E. *The Structure of Science*. N. York & Burlingame: Harcourt, Brace and World, 1961.

Porter, T. *The Rise of Statistical Thinking 1820-1900*. Princeton: Princeton University Press, 1986.

Schultze, Ch. L. "Review of *a Theory of the Consumption Function*." *Science* 127 (1958): 243.

Stigler, S. "The History of Statistics in 1933." *Statistical Science* 11 (1996): 244-52.

———. "Some Correspondence on Methodology between Milton Friedman and Edwin B. Wilson." *Journal of Economic Literature* 32 (1994): 1197-203.

Teira, D., "A positivist tradition in early demand theory." *Journal of Economic Methodology* 13 (2006): 25-47.

———. "Milton Friedman, the Statistical Methodologist.? *History of Political Economy* 39 (2007): 511-28.

Wallis, W. A., and M. Friedman. "The Empirical Derivation of Indifference Functions." In *Studies in Mathematical Economics and Econometrics*, edited by O. Lange, F. McYntire, Th. Yntema, 175-89. Chicago: University of Chicago Press, 1942.