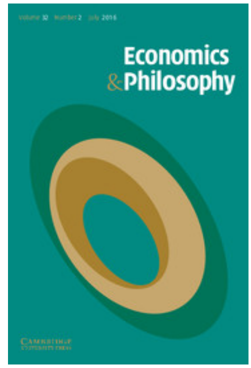

Underdetermination in Economics. The Duhem-Quine Thesis

K. R. Sawyer, Clive Beed and H. Sankey

Economics and Philosophy / Volume 13 / Issue 01 / April 1997, pp 1 - 23
DOI: 10.1017/S0266267100004272, Published online: 05 December 2008

Economics and Philosophy, 13 (1997), 1–23 Copyright © Cambridge University Press

UNDERDETERMINATION IN ECONOMICS. THE DUHEM–QUINE THESIS

K. R. SAWYER
University of Melbourne
CLIVE BEED
University of Melbourne (Retired)
H. SANKEY
University of Melbourne

'Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body'. W. Quine, 1953

1. INTRODUCTION

This paper considers the relevance of the Duhem–Quine thesis in economics. In the introductory discussion which follows, the meaning of the thesis and a brief history of its development are detailed. The purpose of the paper is to discuss the effects of the thesis in four specific and diverse theories in economics, and to illustrate the dependence of testing the theories on a set of auxiliary hypotheses. A general taxonomy of auxiliary hypotheses is provided to demonstrate the confounding of auxiliary hypotheses with the testing of economic theory.

The history of science commonly focuses on the role of certain crucial experiments or observations. The notion of a crucial experiment dates from the seventeenth century in Bacon's *instantiae crucis*, in which the truth or falsity of scientific theories is unerringly revealed by a unique experiment (see Laudan, 1965). Crucial experiments became the dogma of nineteenth century experimental research, especially in optics, thermodynamics and chemistry. While refined versions of a scientific theory

The detailed comments of two anonymous referees and of the editors are gratefully acknowledged.

are often attributed to the results of such crucial experiments, it is rather uncommon, however, for a scientific theory to be completely refuted by a single crucial experiment. Even in these rare cases, the importance of the experiment is typically not recognised immediately. For example, the significance of the Michelson–Morley experiment in theoretical physics was apparently not perceived until Einstein developed his photo-electric theory twenty-five years later (see Lakatos, 1978). It is also the case that, even retrospectively, the outcomes of crucial experiments are seldom regarded as unambiguous. In the social sciences, the possibility of crucial experiments is weakened by what Mill (1984, p. 59) describes as 'the immense multitude of the influencing circumstances, and our very scanty means of varying the experiment'.

Typically, numerous indeterminacies beset the relation between theory and data, thereby undermining the conclusiveness of crucial experiments and more generally, tests of hypotheses. There are three principal sources of indeterminacy. One is observational equivalence. Observationally equivalent hypotheses are indistinguishable from the point of view of observations as noted by Haavelmo (1944). It is impossible for statistical inference to decide between observationally equivalent hypotheses. A confirmatory test result in this case confirms not only the given theory, but also its observationally equivalent alternatives. The problem of identification is that of specifying econometric models in such a way that the problem of observational equivalence does not arise.

A second source of indeterminacy turns on the relation of test to theory. A theory can usually only be tested in the presence of other hypotheses, a set of *auxiliary* hypotheses. The confounding of theory and its auxiliary hypotheses was first identified by a French physicist Pierre Duhem in 1906. This confounding is structurally similar to a third source of indeterminacy, the confounding of several substantial hypotheses belonging to the same theory. Both types of confounding are important in economics; Mongin (1988) illustrates the importance of the third type of indeterminacy in expected utility theory. Discussing methodology in physics, Duhem (1906, p. 187) held that 'the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses'. Confounding is the phenomenon whereby if the experimental results do not agree with the predictions, 'the experiment does not designate which one (hypothesis) should be changed, for it is the whole theoretical scaffolding used by the physicist that is called into question'. (Duhem, 1906, p. 185).

Duhem's thesis remained largely unknown to English readers until its translation in 1954. However, it was acknowledged and reasserted in a different context by the philosopher Willard Quine in 1953. Quine (1953, p. 41, para. 1) in his essay 'Two Dogmas of Empiricism' referred briefly to Duhem's work in his discussion of reductionism. Quine's contention was that hypotheses about the external world could not be tested individually but only as part of a collective set, which, as Cross (1984, p. 83) has observed, is an holistic version of the Duhem thesis. Quine restated Duhem's hypothesis in pure logic.

The Duhem and Quine propositions came to be known as the Duhem–Quine thesis, a term first used by Grünbaum (1963).[1] The Duhem–Quine thesis maintains that theories can be submitted to test only in conjunction with a set of assumptions and rules of inference. These constitute the auxiliary hypotheses and include the simplifying assumptions of the theory, the calibration of the experiment and the axioms of statistical inference used. Cross (1982) was one of the first authors to consider the Duhem–Quine thesis in economics, focussing particularly on the statistical assumptions.

For Duhem and Quine, a negative observational result shows not that the theory is false, but that the complex of theory and auxiliary hypotheses, taken conjointly, is false. *Modus tollens* arguments[2] are thereby weakened as instruments of criticism. This is independent of two other problems in experimental confirmation, the problem of induction and the fallacy of affirming the consequent.

Both Duhem and Quine discussed indeterminacy in a falsificationist rather than a verificationist context. It is difficult to conceive any application of the Duhem problem to verificationism, for as Popper (1959) showed, a singular statement cannot verify a scientific law, while Hempel (1965) demonstrated that the search for verifying instances leads to paradoxes. The question of *which* among several scientific sentences is verified by a positive observational result, which would be the Duhem question in a verificationist context, appears to be meaningless.

According to Duhem and Quine, it is possible to falsify only collective sets of theories, including auxiliary hypotheses, and perhaps encompassing background knowledge. This constraint echoes Lakatos' sophisticated falsificationism, as distinct from naive falsificationism. In Lakatos' terminology, sophisticated falsificationism is concerned with appraising '*sets of theories*', rather than 'an isolated *theory*'; 'the concept of *theory* as the basic concept of the logic of discovery [is replaced] by the concept of *series of theories*'. (1978, pp. 34, 46; original emphasis). Although Lakatos did not use the term 'theory' identically with Duhem or Quine, the essence of Lakatos' claims is not dissimilar from those of Duhem and Quine.

Formally, consider an hypothesis H, an observation statement O and a set A of auxiliary hypotheses $A_1, \ldots, A_n$. If we assume that $H \rightarrow O$,

[1] The evidence that Grünbaum first used the term emerges from private correspondence between Quine and Grünbaum published in Harding (1976, p. 132).

[2] *Modus tollens* is represented by the following schema: if, from a given hypothesis H, we predict an observation O, then if the prediction is false, then so must be the hypothesis H.

then falsification can be represented by a *modus tollens* argument of the form:

$$[(H \rightarrow O) \bullet \sim O] \rightarrow \sim H \tag{1}$$

where $\sim O$ means that O is not observed, and $\bullet$ represents a conjunction of hypotheses. The Duhem–Quine thesis asserts that such arguments rarely, if ever, occur in scientific enquiry. Instead $H \rightarrow O$ is replaced by $H \bullet A_1 \bullet \ldots\ldots \bullet A_n \rightarrow O$, and the falsification $\sim O$ only infers that the antecedent conjunction $H \bullet A_1 \bullet \ldots\ldots \bullet A_n$ is false. Nothing follows about the falsity of H.

The Duhem–Quine thesis admits both stronger and weaker forms. The stronger and weaker forms of the thesis are discussed by Harding (1976, p. xvi) and Laudan (1965, p. 298). Briefly stated they assert:
1. (Stronger): The burden of proof for those who refuse to reject H is to show the existence of auxiliary hypotheses A which would make H compatible with O; that is, for which $H \bullet A \rightarrow$ A.
2. (Weaker): The burden of proof for those who reject H is to show that there are no auxiliary hypotheses A which would make H compatible with $\sim O$; that is, for which $H \bullet A \rightarrow \sim O$. The emergence of stronger and weaker versions of the thesis is attributable to the Grünbaum (1960) critique of Duhem's work. It is generally accepted that the weaker form represents the Duhem–Quine thesis with the greatest fidelity, and indeed Harding (1976) and Laudan (1965) contend that the stronger thesis would not have been endorsed by Duhem.

Implicit in the Duhem–Quine thesis, there are three further ingredients:
1. The pure Duhem–Quine thesis assumes H to be an isolated hypothesis, that is, one that cannot be expressed as the conjunction of several hypotheses. In particular, the thesis is uninformative as to the falsification of systems of hypotheses (such as the geometry of space[3] considered by Grünbaum (1960)), or of science as a whole including the laws of logic. Quine's contention was that only science as a whole was empirically testable; Duhemian conventionalism admitted the possibility that systems of hypotheses such as Euclidean geometry were also testable. Any system of hypotheses subjected to test, however, must be a relevant clustering. As observed by Cross (1984, p. 83), this would surely exclude the conjunctive testing of economic theories and theories of gravitational attraction. One mechanism for determining relevant clusters is the concept of the Lakatos research programme (Cross, 1982), which specifies

[3] By geometry of space, Grünbaum refers to the system of metric relations exhibited by solid bodies, including the specification of terms such as length, direction and congruence. An example of a metric geometry is Euclidean geometry, which can be regarded as the non-unique conjunction of subsidiary hypotheses; for example, absolute geometry and the Euclidean parallel postulate.

groupings of hypotheses as linked together to define a research programme. Duhem did not emphasise the third source of indeterminacy mentioned above, the confounding of several substantial hypotheses belonging to the same theory. However, the logical structure of this indeterminacy is the same as for that of the Duhem–Quine thesis.

2. The set of auxiliary hypotheses will always subsume a substantial component of *background knowledge*. This includes, for example, the laws of logic, but in economics also pertains to such concepts as rationality, market information, market structure and divisibility of commodities. Hausman (1984, pp. 347–8) discusses some of these in the context of general equilibrium models. Popper (1959) extensively discussed the role of background knowledge, and recommended that the various hypotheses and assumptions (whether auxiliary or not) constituting a theory should be tested independently of each other. Interestingly, this maxim does not have a purely experimental flavour. It also leads to the recommendation of axiomatizing scientific theories, since only after completing this task can a scientific statement be checked to be logically independent of another.

3. A set of non-trivial auxiliary hypotheses A will have probability less than one in any empirical setting. Hence, the falsity of the conjunction of hypotheses $H \bullet A_1 \bullet \ldots\ldots \bullet A_n$ will not imply the falsification of H.

One response to the Duhem–Quine thesis, implicit in many contemporary developments in hypothesis testing, is to attenuate the confounding introduced by auxiliary hypotheses. This can be achieved by a number of mechanisms, *inter alia*:

3.1. Choosing auxiliary hypotheses which have a higher empirical probability.

3.2. Specifying only a small number of auxiliary hypotheses, which are necessarily quite general.

3.3. Using hypotheses tests of H which are relatively invariant (or robust) to auxiliary hypotheses.

Many of these mechanisms are used in hypothesis testing in economics. However, reference to the Duhem–Quine thesis in economics has been rather scarce, at least relative to considerations of observational equivalence and underidentification. Some notable exceptions include the formal examination of the thesis by Cross (1982), further developed in Cross (1984), and the classification of auxiliary hypotheses by Hausman (1984, pp. 347–8).

There are compelling reasons for considering Duhem–Quine anew in economics. First, many recent contributions in econometrics, especially the development of robust tests of hypotheses, represent a response to the Duhem–Quine thesis. Robust tests are designed to diminish the influence of specific auxiliary hypotheses, and hence the confounding identified by Duhem–Quine. Secondly, the models of hypotheses which emerge from economic theory are the product of many decisions. An econometric model involves: (1) decisions about how to measure variables, resulting in an unknown amount of measurement error; (2) decisions about how to specify the model, which may involve specification error; (3) decisions to omit the many small causes which influence the dependent variables in the model, which results in the presence of random disturbances (sometimes known as omitted variable error); and (4) decisions about how to calculate the estimates of the model, which involves approximation error. (The estimate of the square root of 2 involves approximation error.) Economists construct a probability model for these four types of errors. One of the hallmarks of contemporary econometrics is the emphasis on error diagnosis, that is the testing of subsets of auxiliary hypotheses. These tests are attenuated by problems of the Duhem–Quine type arising from other auxiliary hypotheses. For example, tests for the temporal correlation of errors will usually be conditional on some assumed distribution of the errors such as normality. An hierarchical Duhem–Quine problem thereby emerges, with sets of auxiliary hypotheses being tested sequentially for their effect on the test of the principal hypothesis.

The purpose of the present paper is to consider the Duhem–Quine thesis in economics; in particular to identify and classify variants of auxiliary hypotheses and to amplify their effects on hypothesis testing. In Section 2, we explore several diverse theories in economics and their attendant auxiliary hypotheses. Section 3 provides a taxonomy of auxiliary hypotheses which extends and generalizes the auxilary hypotheses of Section 2. In Section 4 we discuss the confounding of auxiliary hypotheses with tests of economic theory.

2. DUHEM–QUINE IN THE REALM OF ECONOMICS

Economic analysis usually begins with economic laws. It is commonly asserted these laws were established inductively by psychology or to be technical truths established in the natural sciences (Hausman, 1984, p. 350). The principle of non-satiety in consumer theory is a form of economic law as is the law of diminshing marginal utility. A precise statement of a collection of economic laws with appropriate characterizations of institutional and informational background leads to the development of an economic theory. This is the deductive approach to economics, propounded by John Stuart Mill (1836) and elaborated by Hausman (1989, p. 116). One of the most significant examples of an economic theory is equilibrium theory, a formal statement of which is given in Hausman (1984, p. 345). Economic theories assume a substantial component of background knowledge of varying levels of refinement. For example, equilibrium theory can be further refined to a general equilibrium theory by the attachment of auxiliary hypotheses relating to market structure, commodity divisibility and interdependence. This is the formalization implicit in Debreu's *Theory Value of Value* (1959) and is well discussed in Hausman (1984). Economic theories are often also translated into testable hypotheses, through the conjunction of various statistical hypotheses including the laws of statistical inference.

Testing an economic theory tests:

1. A set of economic laws.
2. A composition of background knowledge
3. A set of auxiliary economic hypotheses
4. A set of auxiliary statistical hypotheses.

Background knowledge, economic and statistical assumptions represent the auxiliary hypotheses of the Duhem–Quine thesis. Background knowledge includes the laws of logic and inference, the antecedent economic and statistical theory, and the terminology and symbolism used. There is no possibility that the confounding effects of background knowledge can be diversified away, so that in this sense it represents the most fundamental problem of the Duhem–Quine type. It is also the case that it cannot be classified; for this reason it is not included in the discussion below.

No economic theory can ever be regarded as strictly true, but rather as an approximation to observed phenomena. One method for characterizing this approximation is to replace an hypothesis H by a conjunction $H \bullet \varepsilon$ where ε imposes assumptions about the distribution of errors. This of course alters the form of the Duhem–Quine thesis, since there will be some confounding between ε and any auxiliary hypotheses which make $H \bullet \varepsilon$ compatible with $\sim O$. For example, the weaker version of the thesis now requires those who reject H to show that for a wide class of approximate hypotheses $H \bullet \varepsilon$, there is no set of auxiliary hypotheses *A* which would make $H \bullet \varepsilon$ compatible with $\sim O$. A priori, the hypothesis ε is typically unknown, so that falsification of $H \bullet \varepsilon$ is further weakened by uncertainty in ε. In the following discussion of auxiliary hypotheses, neither background knowledge nor approximation error are referenced because of problems in classification and the uncertainty of their specification.

In this essay, we exposit the Duhem–Quine thesis by considering four economic theories and the auxiliary hypotheses inserted to test them. The purpose of this exposition is to show the evolution of a testable hypothesis from an economic theory, and, in particular, to illustrate the diversity of auxiliary hypotheses. This diversity is evident across different economic theories, but also for a given economic theory. A given theory then generates a diverse family of conjunctions of the theory and auxiliary hypotheses. The Duhem–Quine thesis implies that only these conjunctions can be falsified, and not the underlying theory.

The theories are:

Theory 1: (Positive duration dependence)
Unemployed individuals are more likely to accept wage offers the longer they have been unemployed.

Theory 2: (Phillips curve)
Lower unemployment may be purchased with higher inflation.

Theory 3: (Neutrality or policy ineffectiveness)
Policy, and expected monetary changes in particular, have no effect on long run economic growth.

Theory 4: (Arbitrage pricing)
Expected asset returns are a linear combination of a finite and constant number of risk premia, so that no super-normal returns may be earned.

These theories were selected from a plethora of possibilities to represent the richness of economic theories and the auxiliary hypotheses which support them. Mongin (1988), for example, provides an insightful discussion of expected utility theory, and in particular the importance of the underlying axioms of independence and the lottery principle. Each of the theories selected in this section has conceptually different antecedents; for example, the Phillips curve was originally an empirical proposition, while the arbitrage pricing theory was developed in a theoretical setting in response to the empirical limitations of the capital asset pricing model.

Consider *Theory 1.* It admits of at least two formalizations. One simple formalization is the job search model with the property that an individual accepts employment if the offered wage exceeds the reservation wage. Positive duration dependence is said to exist if the reservation wage declines during the job search period. When reservation wages are observed, positive duration dependence can then be tested directly. The job search model assumes, *inter alia,* that:

1. Job offers arrive as realizations of a Poisson process with mean λ_t.
2. The distribution of wage offers is known or can be learned.
3. The cost of search is non-zero (C_t per period), and may be constrained over time.

Under these assumptions, the reservation wage (W_t^*) in period t satisfies

$$W_t^* = -C_t + \frac{\lambda_t}{r} \int_{W_t^*}^{\infty} (w - W_t^*) dF_e(w) \qquad (2)$$

where dF_e is the assumed wage offer distribution, and r is a discount factor (see Mortenson, 1986 or Sawyer, 1989 for an elaboration). When W_t^* is observed, the existence and cause of positive duration dependence can be assessed by estimating equation (2). The formalization in this case introduces three auxiliary hypotheses related to the distribution of job offers, the distribution of wage offers and the cost of job search.

When reservation wages are not observed, it is simpler to model the probability of exit from unemployment directly. In this case, we have

$$P_t = h(X_t, \beta f(\theta) \tag{3}$$

where P_t is the probability of exit from unemployment, h is a prescribed baseline hazard function,[4] X is a set of covariates with sensitivities β, and f(θ) is a heterogeneity term[5] to represent attributes not included in X. There are (at least) three auxiliary hypotheses in this formalization pertaining to the separation of the baseline hazard function and the heterogeneity term, the specification of the baseline hazard function and the selection of the attributes X.

The two formalizations (2) and (3) admit two different sets of auxiliary hypotheses; the choice of formalization depends first on the availability of data on reservation wages but secondly on the plausibility of auxiliary hypotheses – for example, the plausibility of the given wage offer distribution dF_e (w) vis-à-vis the baseline hazard function $h(X_t, \beta)$. Perforce, tests of duration dependence are conditioned by these auxiliary hypotheses. It is possible, for example, for the two formalizations to yield irreconcilable conclusions simply because the wage offer distribution and the hazard function are incompatible. This exemplifies the stronger version of the Duhem–Quine thesis; essentially that while different conjunctions of hypotheses are falsifiable, the truth or falsity of the fundamental hypothesis of positive duration dependence remains uncertain.

Let us now consider *Theory 2*, the trade-off between output and inflation. This can be formalized and tested in many ways. The literature on the standard Phillips curve alone is quite voluminous (see Desai, 1984 and Wulwick, 1987). Much of the discussion is centred on whether the relationship between the rate of inflation and unemployment applies only in the short-run, or whether there is a long-run trade-off. A predominant view accepts that there is a negative association between unemployment and inflation in the short-run, although proponents of the natural rate hypothesis in particular would argue against any causal relation. The long-run relation is less well-defined; a conventional view is that, at least for relatively low rates of inflation, the slope of the curve drawn in the

[4] A hazard function is a function representing the probability of switching from one state (unemployment) to another.

[5] Heterogeneity refers to attributes of individuals which may not be measured by concomitants. A common example is individual ability.

long-run inflation-output space is zero. There is no possibility of a long-run trade-off.

There have been many formalizations and tests of the relation between unemployment and inflation. We have selected three formalizations to amplify three subsidiary questions: first whether the relationship applies to voluntary or involuntary unemployment, secondly whether the relationship is causal, and thirdly whether the relationship extends to sub-markets. Each of these questions has a long pedigree in the Phillips curve (and its variants) literature. Tobin (1972, p. 4) suggests an interesting mechanism for testing for involuntary unemployment only. He posits the question whether labour would be willing to accept an offer of a job at real wage rates lower than the prevailing wage rate in order to induce lower unemployment. An affirmative answer to this contrived question indicates voluntary unemployment and an output-inflation trade-off consistent with neoclassical job search theory. To test this question requires a controlled labour market mechanism, in particular the ability to reduce real wages. As a consequence, the auxiliary hypotheses in this case include variables such as the bargaining power of employers and unions.

An alternative explanation of the output-inflation trade-off is that labour markets are in disequilibrium and the consequent involuntary unemployment induces some real wage adjustment. This hypothesis of the output-inflation trade-off, which is distinct from job search theory, is tested by Nevile (1979) in a paper with the evocative title 'How voluntary is unemployment? Two views of the Phillips curve'. Nevile uses a Granger causality test to assess the direction of association between movements in wages and changes in excess demand. The implicit auxiliary hypothesis is that actual wage adjustments are captured by the metronomics imposed; for example, if wages adjust within one month to excess demand in the labour market, then quarterly data would not be optimal for assessing the direction of association. But an additional hypothesis is also imposed – that non-causality can be fully tested using the Granger method.[6]

A third approach to the testing of the Phillips curve is to consider the level of aggregation required to induce some output-inflation trade-off. It is not apodictic that a trade-off should always exist at an economy-wide level. Indeed, the micro-theoretic foundations of the Phillips curve (see for example Holt, 1970) do not preclude the possibility of spillovers between sectors of the economy and these spillovers may annul Phillips curve effects at the aggregate level. A number of authors have attempted to identify Phillips curve effects in disaggregated markets; for example, Sawyer and Alauddin (1989) used industry wage and unemployment

[6] Causality is represented by temporal causality. For a discussion see Sims (1972).

data in the Australian economy. Testing the Phillips curve at the macro level involves testing the micro-theoretic foundations and the unspecified process of aggregation across labour markets; testing at the micro level tests the micro-theoretic foundations and the assumption that labour markets can be segmented. Important auxiliary hypotheses relating to aggregation and segmentation are then introduced, simply by the choice of labour market to be tested.

A review of the Phillips curve literature then suggests some rich variants of auxiliary hypotheses, simply because of the way the output-inflation trade-off is examined. These auxiliary hypotheses supplement and even supplant other auxiliary hypotheses relating to the choice of concomitant variables such as price expectations. Some auxiliary hypotheses may be more important than others in the testing of the output-inflation trade-off; the representation of price expectations may be characterized as a substantial hypothesis and hence the third type of indeterminancy identified in Section 1 may be relevant. However, the Duhem–Quine thesis applies to all conjunctions of the fundamental and auxiliary hypotheses, regardless of their importance. Hence, for example, the Phillips curve cannot be falsified independently of some assertion as to the formation of price expectations.

An important proposition in economics, *Theory 3*, is the long-term neutrality of policy decisions, especially those relating to anticipated changes in money. Under neutrality, expected changes in money have no effect on long-run real economic growth. To test this proposition, there have been at least two paradigms. Barro (1978) embedded both anticipated and unanticipated monetary changes in a Lucas–Sargent–Wallace (LSW) aggregate supply curve and tested for the effect of the anticipated changes on real output. Formally, Barro specifies the supply curve as:

$$Dy_t = A(L)\,DM_t^u + B(L)\,DM_t^e + v_t \tag{4}$$

where Dy_t is the deviation from trend in real output in period t, DM_t^u and DM_t^e are the unanticipated and anticipated changes in monetary growth in period t, and A(L), B(L) are lag polynomials[7] representing the responses of output to contemporaneous and lagged changes in monetary growth. The neutrality or policy ineffectiveness proposition implies all coefficients in B(L) are zero.

An alternative paradigm, a corollary of the developments in non-stationary time series in the 1980s, is to consider the Wold decomposition of real output. For ease of exposition, we reference the simplest decomposition, one without either a drift term or higher orders of lags.[8]

[7] The lag operator L is that which transforms a variable to its value in the previous period, so that $Ly_t = y_{t-1}$.

[8] A drift term allows the process to have non-zero deterministic growth.

$$y_t = \lambda y_{t-1} + \varepsilon_t \qquad (5)$$

where y_t is real output in period t, ε_t is an error term assumed white noise.[9]

Noting that (5) can be written as an infinite moving average of the innovations ε

$$y_t = \varepsilon_t + \lambda \varepsilon_{t-1} + \lambda^2 \varepsilon_{t-2} + \dots \qquad (6)$$

a test for neutrality reduces to a test that λ equals unity in the autoregressive process (5). For, if λ equals unity, we find that past innovations ε_{t-s}, $s > 0$, have equal influence on real output y_t. Thus the effect of shocks to aggregate supply will persist through time, rendering policy decisions designed to stabilize such shocks ineffective. The test that λ equals unity is called the unit root test.[10]

Policy ineffectiveness (neutrality) will then be implied if either the anticipated changes in money in the LSW model have no effect on real output or if past innovations in real output persist over time. Unsurprisingly, the Barro and unit root approaches assume different auxiliary hypotheses. Barro assumes a LSW supply curve which is well-specified; that is, deviations in real output, anticipated and unanticipated monetary changes are all measured correctly, and the structural decomposition of real output into contemporaneous and lagged DM^u and DM^e is without error. For the unit root approach, the Wold decomposition is assumed correct, so that second and higher order lags in y_t have negligible effect on the present value of output. The auxiliary hypothesis of both approaches is therefore the absence of specification error, although the probable error assumes a different form in each case. The Duhem–Quine thesis asserts that policy ineffectiveness cannot be refuted in disjunction from this hypothesis pertaining to specification error.

As a final proposition, ***Theory 4***, we consider a contemporary theory in finance, the arbitrage pricing theory (APT) which asserts that the expected return on any asset is a linear combination of a small number of economy-wide risk premia. APT was initiated by Ross (1976) and has assumed two forms:

(1) The original structure proposed by Ross which uses a strict factor structure for the generating process of asset returns.
(2) A variant suggested by Chamberlain and Rothschild (1983) which uses an approximate factor structure for the generating process of asset returns. Both strict and approximate factor structures assume

[9] White noise refers to an error process with zero mean, uncorrelated across observations and with finite, constant variance.

[10] A further elaboration of the motivation for the testing of a unit root is given in Stock and Watson (1988).

that in a finite economy each asset return is determined by a small set of macroeconomic factors. These factors include, for example, unanticipated changes in inflation or in the term structure of interest rates. In the strict factor structure, correlations between assets are channelled entirely through the macroeconomic factors while in the approximate factor structure some additional correlation is permitted – for example, due to industry effects.

For both types of returning generating processes (strict and approximate factor structures), it has been shown that in the absence of arbitrage opportunities,[11] expected asset returns are approximately equal to a linear combination of a small number of risk premia which are correlated with unanticipated changes in the macroeconomy. Curiously, the empirical literature in finance has not focussed on this main implication of APT, namely the determination of the set of risk premia associated with a given asset. Instead, most of the empiricism has been concerned with the identification of the returns generating process, either the strict factor structure (for example Roll and Ross, 1980) or the approximate factor structure (Trzincka, 1986).

This testing problem is an example where the auxiliary hypothesis has actually become the hypothesis under test. That is, the main hypothesis is the relationship between expected returns and risk premia, the auxiliary hypothesis is the returns generating process which produces the expected returns and identifies a factor structure; yet the auxiliary hypothesis which is a necessary antecedent to arbitrage pricing has become the central tenet of empirical testing.

In this tale of four theories in economics and finance, we have noted in each case a number of variants of the theory. Each of these different forms connotes different auxiliary hypotheses. In some cases, such as APT, the auxiliary hypothesis actually dominates the main hypothesis; that is, it supplants the central hypothesis that expected returns are linearly related to risk premia. Typically though, auxiliary hypotheses interfere to a lesser extent in the testing methodology. But in all cases there must be some confounding due to the presence of the auxiliary hypotheses; for example, it is possible to conclude that Phillips curve effects are present in the economy when using a disaggregated approach but conversely to deny their effects when empirically testing at the aggregate level. As contended by Duhem and Quine, it is therefore not possible to irrevocably refute any of the *Theories 1* to *4* as isolated hypotheses, but only to refute a potentially infinite set of conjunctions of hypotheses.

[11] An arbitrage opportunity exists when a costless net investment can earn a positive return with zero risk.

The four theories we have considered illustrate the richness of possible auxiliary hypotheses. In Section 3, we classify types of auxiliary hypotheses frequently used in economics. In Section 4, we examine the possible confounding introduced by each approach.

3. A TAXONOMY OF AUXILIARY HYPOTHESES

There are at least two ways of classifying the auxiliary hypotheses of the Duhem–Quine thesis. First, auxiliary hypotheses can be characterized by their generality (or specificity). This is the approach adopted in this section, essentially because it is simple. An alternative taxonomy would be to order auxiliary hypotheses according to their effect on the outcome of the hypothesis test, and this is implied by Section 4.

To formalise these concepts, we introduce the following notation:

1. Let H be the hypothesis under test. Each of the four theories in Section 2 is an example of H.
2. We assert that A is an auxiliary hypothesis for the testing of H, if the outcome of a test T of H is also dependent on A. More formally,

$$T = T(H \bullet A) \tag{7}$$

where T denotes the outcome of the test, and $H \bullet A$ means that both H and A are present. In statistics, A is often characterized as a set of nuisance parameters, that is, parameters which are not of primary interest.

Consider two auxiliary hypotheses A_1 and A_2 and suppose they generate test outcomes

$$\begin{aligned} T_1 &= T(H \bullet A_1) \\ T_1 &= T(H \bullet A_2) \end{aligned} \tag{8}$$

A_1 and A_2 may be classified directly according to their levels of generality; A_1 is more general than A_2 if variations in A_1 induce greater changes in the method of testing than induced by variations in A_2. We identify five types of auxiliary hypothesis:

Type 1: The most general type of auxiliary hypothesis is the general methodological setting. This confers different types of auxiliary hypotheses, depending on the methodology used. The translation of methodologies into statements and auxiliary hypotheses is non-trivial, and certainly not of the same logical form as statistical assumptions concerning, for example, normality. In particular, three methodologies should be distinguished for the testing of economic propositions:

1. Econometric and other stochastic methods including simulation. This

methodology involves testing an econometric model against real-world observations, or simulated observations. The methodology assumes a random data-generating process, and as a consequence emphasizes statistical assumptions. Most of the theories in Section 3 are typically tested econometrically.

2. A number of deterministic methods exist for testing economic theories, including techniques such as comparative statics, deterministic simulation, game theory, linear and nonlinear programming. In this methodology, there is less emphasis on real world observations, and more on theoretical consistency. For example, the arbitrage pricing theory (***Theory 4*** in Section 3) can be assessed using nonlinear programming techniques.
3. Experimental methods. The emergence of experimental economics in recent years signals a new methodology for the testing of economic hypotheses. In this case, there are many auxiliary hypotheses which arise due to the experimental method, for example, the nature of the monetary rewards, the assumption of no strategic interactions between the experimenter and the subjects and so on. Mongin (1988) provides a useful discussion for the problem of testing expected utility theory.

An experimental setting is clearly different from an econometric setting. It is apparent that the outcomes of tests of economic propositions will clearly reflect the difference of the setting. However, it is rare for economists to explore different methodologies within the context of the one problem; thus sensitivity to this most important of auxiliary hypotheses is seldom investigated. An exception is the recent literature in chaotic dynamics, where tests for determinism against stochasticity have been constructed for economic time series (see Brock, 1986 for an example). A further exception, but outside the realm of economics, is a study by Efron and Thisted (1976) who ponder the question: *'How many words did Shakespeare know?'* from both a deterministic and stochastic perspective.

Type 2: Within the methodologies of Type 1, there are a number of subsidiary methodologies which confer different auxiliary hypotheses. There are, for example, a number of variants of econometric methodology. A non-exhaustive list includes:

1. Classical methodology, based on the principle of maximum likelihood estimation with tests of statistical hypotheses (Fisher, 1935; Neyman and Pearson, 1928) covers imposing economic hypotheses which imply restrictions on reduced form parameters. It also includes
   (a) Atheoretical econometrics, where a minimal set of prior informa-

tion is imposed; for example the vector autoregressive methodology used by Sims (1980).
   (b) The error correction and cointegration approach of contemporary macroeconometrics (Engle and Granger, 1987).
   (c) Leamer's sensitivity analysis, where the robustness of certain focus coefficients is analysed (see Leamer, 1977).
2. Non-parametric and semi-parametric estimation, embodying the notion that not all economic information can be channelled through parameters (for example, Horowitz and Neumann, 1987 and Ullah, 1989).
3. Data analysis, which includes the exploratory data analysis of Tukey (1977). An example of economic research that uses data analysis of an exploratory type is Friedman and Schwartz (1963).
4. Bayesian analysis as illustrated by Tsurumi and Tsurumi (1983).[12]

This second type of auxiliary hypothesis, encompassing the form of econometric methodology used, again introduces a dependency, that is, the outcome of a test of an economic proposition is not invariant to the choice of econometric methodology. Typically, only one econometric methodology is used, so that it is difficult to discern the sensitivity of econometric results to the methodology used; a notable exception is the research of Hendry and Ericsson (1991) where the classical methodology of modern econometrics is compared to the data analysis used by Friedman.

Type 3: The third type of auxiliary hypothesis refers to the selection of the sample. In particular, this encompasses:

1. The nature of the data – whether it be time-series, cross-sectional or longitudinal.
2. The metronomics of the data, that is, whether it is daily, weekly, monthly or some other frequency.
3. The selection of the data, whether random sampling, stratified sampling or, in time-series data, episodal sampling. This last sampling technique is used for example in Friedman and Schwartz (1982).
4. The level of aggregation in the data; in cross-sectional data, for example, data may be analysed at the individual level, the household level or aggregate level.

Selection of the sample is an important auxiliary hypothesis which

[12] We note the similarity of some of these variants of econometric methodology. In particular, Leamer's sensitivity analysis is essentially a Bayesian technique, and there is also similarity between non-parametric methods and exploratory data analysis.

confers an immediate dependency on the outcomes of hypothesis tests. For a compelling example of how the results of hypothesis tests change with sample selection, see Rush and Waldo (1988). They illustrate that the presence of additional observations may alter the outcomes of non-nested tests of hypotheses.

Type 4: In some problems, there are auxiliary hypotheses which are essentially *sine qua non* for the testing of the main hypothesis. Two examples are:

1. Rational expectations which can only be tested conditional on a given model specification. Rationality cannot be tested without an implied stochastic model, which allows certain parametric restrictions to be inferred.
2. Arbitrage pricing for which the testing is necessarily predicated on an identified factor structure (see Theory 4 in Section 2). The model specification for testing rationality and the factor structure for arbitrage pricing constitute Type 4 auxiliary hypotheses.

Type 5: The most salient auxiliary hypotheses in the testing of an economic proposition are the statistical assumptions used in the estimation and testing of hypotheses, and commonly referred to as econometric conventions. Cross (1982) discusses these in some detail, but it is also reflected in other reviews of empirical work (e.g. DeWald et al, 1986).

Briefly a single equation econometric specification

$$y = f(X, \beta) + u \tag{9}$$

involves the following assumptions:

1. That u, the data generating process, follows some probability distribution, with a covariance matrix of known form.
2. That X, the set of exogenous variables, is complete and uncorrelated with u.
3. That f, the response function of the data, is well-specified.
4. That there is only one endogenous variable in the system; alternatively if there is more than one endogenous variable, then this endogenous variable is uncorrelated with other endogenous variables in the system.
5. That measurement error is non-systematic.

These auxiliary hypotheses are closely scrutinized by practitioners; a substantial econometric literature is concerned with the sensitivity of econometric estimates and tests to changes in the conventional

Table 1: A Classification Of Auxiliary Hypotheses

THEORY	AUXILIARY HYPOTHESES	
Theory 1 Duration Dependence	Structural model or baseline hazard	**Type 2**
	Economic assumptions (e.g., Poisson process)	**Type 5**
	Econometric assumptions (functional form)	**Type 5**
Theory 2 Phillips Curve	Controlled labour market (Tobin)	**Type 1**
	Micro or aggregate data	**Type 3**
	Rational expectations	**Type 4**
	Concomitant variables (e.g., expectations)	**Type 5**
Theory 3 Neutrality	Unit root test or Barro test	**Type 2**
	Selection of sample	**Type 3**
	Assumption of rationality	**Type 4**
	Absence of specification error	**Type 5**
Theory 4 Arbitrage Pricing	Assumption of factor structure	**Type 4**
	Strict or approximate factor structure	**Type 5**
	Absence of specification error	**Type 5**

assumptions (for example, the effect on ordinary least squares estimates and t-statistics of moving from Gaussian white-noise errors to non-Gaussian white-noise errors, or from white-noise errors to autocorrelated errors; see Maddala, 1977, Chapter 12).

The classification above into five types of auxiliary hypotheses is not designed to be exhaustive, but it will subsume many of the auxiliary hypotheses frequently invoked in testing economic theories. In Table 1, we classify the auxiliary hypotheses introduced in Section 2 in our discussion of four theories in economics.

The above classification is also not unique. Rather, we have emphasized some of the more significant auxiliary hypotheses. It is evident that Type 5 auxiliary hypotheses are present in most problems. More general auxiliary hypotheses are often present, if only implicitly. This completes the discussion of auxiliary hypotheses and their classification. We now consider the effect of such ancillarity on the testing of the main hypothesis.

4. CONFOUNDING OF AUXILIARY HYPOTHESES

The existence of auxiliary hypotheses confounds the testing of the principal hypothesis (H); that is, the results of a test T of H are not invariant to the choice of auxiliary hypotheses. There are two characterizations of this lack of invariance:

1. For two auxiliary hypotheses of the same type, two different

hypothesis test outcomes may be generated. Formally, for two auxiliary hypotheses A_1 and A_2 of the same type, then the respective test outcomes T_1 and T_2 defined in (8) above will differ; that is, the event that T_1 equals T_2 is not certain.

2. For four auxiliary hypotheses A_1 , A_2 , A_3 , A_4 with A_1 and A_2 of type I and A_3 and A_4 of type J, a priori we expect:
   (a) The respective test statistics T_1 ,T_2 ,T_3 and T_4 to differ.
   (b) If I is a more general hypothesis than J, the differences of T_1 and T_2 will be larger than the differences of T_3 and T_4.

The corollary of these axioms is that auxiliary hypotheses affect the outcomes of hypothesis testing; it is possible for example to accept a given hypothesis H under one set of auxiliary hypotheses and reject H under another set of auxiliary hypotheses. Furthermore, this sensitivity to auxiliary hypotheses is likely to be greater for the more general types of auxiliary hypotheses in Type 1 than the more specific types in Type 5; the results from using a deterministic model as opposed to a stochastic model are likely to be vastly different than from changing the error structure in an econometric model.

There are two mechanisms by which empirical economists attempt to assess the effect of auxiliary hypotheses. One method is a sensitivity analysis, where the effect on the main hypothesis is simulated for a number of perturbations of the auxiliary hypotheses. Sensitivity analyses can assume various forms; in general equilibrium modelling, for example, it is common to vary assumptions pertaining either to parameters or exogenous variables in order to assess the impact on certain endogenous variables. Tests about the future time path of endogenous variables are thereby linked directly to auxiliary hypotheses concerning the exogenous variables.

Sensitivity analyses also appear in econometric work; designs of Monte Carlo experiments in econometrics usually include responses to changes in various auxiliary hypotheses such as the error structure. These responses are often formalized as response surfaces (see Ericsson, 1986), For example, when assessing the statistical performance of a test for autocorrelation, it is common to consider how the power varies as the errors change from the normal distribution to an error distribution which is non-normal, usually leptokurtic or heavy-tailed. If a statistical test is invariant to changes in the auxiliary hypotheses, it is deemed a *robust* test. The term *robust* was coined by Box (1953), and is generally regarded as a desirable property of hypothesis tests. The concept of robustness has been extremely influential in theoretical statistics in the last thirty years, but only relatively recently has it found its way into the econometric literature (see, for example, Thursby, 1981; Koenker, 1982; Horowitz and Neumann, 1987).

The second mechanism for assessing the effect of auxiliary hypotheses is to test the auxiliary hypotheses directly. This conforms with Popper's (1959) maxim noted earlier. It has become common practice in econometrics to test for the economic and statistical assumptions which underscore empirical work, so that practitioners can now choose from an array of tests which test:

1. Determinism against stochasticity (Type 1) (Brock, 1986).
2. Rationality (Type 4) (Barro, 1978)
3. Statistical error structure (Type 5). A good discussion of these types of tests is contained in Bera and Jarque (1982).

Significantly though, there are very few tests which discriminate between different types of econometric methodology (Type 2 auxiliary hypotheses) or different sample periods and metronomics (Type 3 auxiliary hypotheses).

Sensitivity analyses and direct tests of auxiliary hypotheses have contributed to our understanding of how auxiliary hypotheses affect the results of hypothesis tests. It is now more usual in applied economic work for researchers to recognize the likely effects of changes in economic and statistical assumptions. Such cognizance can only enrich the quality of succeeding economic models and attenuate the effects of the Duhem–Quine thesis.

The effects of the Duhem–Quine thesis cannot, of course, be completely diversified away. As noted in the discussion in Section 2, in all tests of economic hypotheses there is a substantial component of *background knowledge* for which there can be no properly defined sensitivity analysis, nor formal testing. For auxiliary hypotheses of the five types identified in Section 4, sensitivity analyses and tests of auxiliary hypotheses exist, but they are conditioned by the range of alternative auxiliary hypotheses considered. It is self-evident that no hypothesis test can be made robust to the set of possibly infinite perturbations in auxiliary hypotheses. Finally, tests of auxiliary hypotheses themselves are affected by problems of the Duhem–Quine type, so that we cannot separately determine their plausibility. The presence of background knowledge, the impossibility of a universal robust test, and the weakness of auxiliary hypothesis tests all infer that the confounding effects implied by the Duhem–Quine thesis can only be marginalized, and never fully eliminated.

5. CONCLUDING DISCUSSION

In its pure form, the Duhem–Quine thesis suggests that no analytical hypothesis can be disentangled from a series of supporting or auxiliary hypotheses which permit its testing. These auxiliary hypotheses will always include a substantial component of background knowledge, which cannot be diversified away. In economics, the testing of hypotheses is further complicated by the approximate nature of theoretical hypotheses. The error in approximation constitutes an auxiliary hypothesis of typically unknown dimension.

w>In the present paper, the Duhem–Quine thesis has been illustrated for a number of propositions in economics. Types of auxiliary hypotheses have been classified and their confounding with hypotheses tests discussed. For auxiliary hypotheses which are quite specific, such as Type 5, the effects of Duhem–Quine can be lessened by adopting tests which are robust to variations in the auxiliary hypotheses. However, for more general auxiliary hypotheses such as the choice of methodological setting, the problem of making hypothesis tests robust to this choice is of far greater difficulty. The challenge of future work in economics is to continue to identify auxiliary hypotheses, to recognise and even quantify their effects, and finally to diminish the effects of the Duhem–Quine thesis by constructing tests robust to many types of auxiliary hypothesis.

REFERENCES

Barro, R. J. 1978. 'Unanticipated money, output, and the price level in the United States'. *Journal of Political Economy*, 86:549–80

Bera, A. K. and C. M. Jarque, 1982. 'Model specification tests: a simultaneous approach'. *Journal of Econometrics*, 20:59–82

Box, G. E. P. 1953. 'Non-normality and tests on variances'. *Biometrika*, 40:318–35

Brock, W. A. 1986. 'Distinguishing random and deterministic systems' *Journal of Economic Theory*, 40:168–95

Chamberlain, G. and M. Rothschild. 1983. 'Arbitrage, factor structure and mean-variance analysis on large asset markets'. *Econometrica*, 51:1281–304

Cross, R. 1982. 'The Duhem–Quine thesis, Lakatos and the appraisal of theories in macroeconomics'. *Economic Journal*, 92:320–40

Cross, R. 1984. 'Monetarism and Duhem's thesis' In *Economics in Disarray*. P. Wiles and G. Routh (eds.). pp. 78–99. Basil Blackwell

Desai, M. 1984. 'Wages, prices and unemployment a quarter century after the Phillips curve'. In *Econometrics and Quantitative Economics*. D. F. Hendry and K. F. Wallis (eds.). pp. 253–73. Basil Blackwell

DeWald, W. G., J. Thursby and R. Anderson. 1986. 'Replication in empirical economics: the journal of money, credit and banking project'. *American Economic Review*, 76:587–603

Duhem, P. 1906. *La Theorie Physique, Son Object, Sa Structure*. Paris, 2nd revised ed. 1914, reprinted by Librairie Vrin, Paris, 1981. The version used here is *The Aim and Structure of Physical Theory*. 1954. Trans. P. Wiener. Princeton University Press

Efron, B. and R. Thisted. 1976. 'Estimating the number of unseen species: how many words did Shakespeare know?' *Biometrika*, 63:435–47

Engle, R and C. Granger. 1987. 'Co-integration and error correction: representation and testing'. *Econometrica*, 55:251–76

Ericsson, N. 1986. 'Post-simulation analysis of Monte-Carlo experiments: interpreting Pesaran's (1974) study of non-nested hypothesis test statistics'. *Review of Economic Studies*, 53:691–707

Fisher, R. A. 1935. 'The logic of inductive inference'. *Journal of the Royal Statistical Society*, 98:39–82

Friedman, M. and A. Schwartz. 1963. 'Money and business cycles'. *Review of Economics and Statistics*, 45:S32–78

Friedman, M. and A. Schwartz. 1982. *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices and Interest Rates, 1867–1975*. University of Chicago Press

Grünbaum, A. 1960. 'The Duhemian argument'. *Philosophy of Science*, 27:75–87

Haavelmo, T. 1944. 'The probability approach in econometrics'. *Econometrica*, 12 Supp: 1–115

Harding, S. G. (ed.) 1976. *Can Theories Be Refuted?* Reidel

Hausman, D. 1984. 'Are general equilibrium theories explanatory?' In *The Philosophy of Economics: An Anthology*. D. Hausman (ed.). pp. 344–59. Cambridge University Press

Hausman, D. 1989. 'Economic methodology in a nutshell'. *Journal of Economic Perspectives*, 3:115–27

Hempel, C. G. 1965. *Aspects of Scientific Explanation*. Free Press

Hendry, D. and N. Ericsson. 1991. 'An econometric appraisal of UK money demand'. In *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna Schwartz. *American Economic Review*, 81:8–38

Holt, C. 1970. 'Job search, Phillips' wage relation and union influence: theory and evidence'. In *Microfoundations of Employment and Inflation Theory*, E. Phelps et al, (eds.). pp. 52–123. Norton

Horowitz, J. and G. Neumann. 1987. 'Semiparametric estimation of employment duration models'. *Econometric Reviews*, 6:5–40

Koenker, R. 1982. 'Robust methods in econometrics'. *Econometric Reviews*, 1:213–55

Lakatos, I. 1978. *The Methodology of Scientific Research Programmes*. Cambridge University Press

Laudan, L. 1965. 'Grünbaum on the Duhemian argument' *Philosophy of Science*, 32: 295–9

Leamer, E. E. 1978. *Specification Searches: Ad Hoc Inference with Non-Experimental Data*. Wiley

Maddala, G. S. 1977. *Econometrics*. McGraw Hill

Mill, J. S. 1936. 'On the definition and method of political economy'. In *The Philosophy of Economics: An Anthology*. D. S. Hausman (ed.). pp. 52–69. Cambridge University Press

Mongin, P. 1988. 'Problèmes de Duhem en théorie de l'utilité espérée'. *Fundamenta Scientiae*, 9:299–327

Mortensen, D. T. 1986. 'Job search and labor market analysis'. In *Handbook of Labor Economics*. Vol. 2. O. Ashenfelter and R. Layard (eds.). pp. 849–919. North-Holland

Nevile, J. 1979. 'How voluntary is unemployment? Two views of the Phillips curve'. *Journal of Post Keynesian Economics*, 2:110–19

Neyman, J and E. S. Pearson. 1928. 'On the use and interpretation of certain test criteria for purposes of statistical inference'. *Biometrika*, 20a, Part1:175–240; Part2:263–94

Popper, K. 1959. *The Logic of Scientific Discovery*. Hutchinson

Quine, W. O. 1953. 'Two dogmas of empiricism'. In *From a Logical Point of View*. Harvard University Press. First published in *The Philosophical Review*: 60

Roll, R. and S. Ross. 1980. 'An empirical investigation of the arbitrage pricing theory'. *Journal of Finance*, 35:1073–103

Ross, S. 1976. 'The arbitrage theory of capital asset pricing'. *Journal of Economic Theory*, 13:341–60

Rush, M. and D. Waldo. 1988. 'On the policy ineffectiveness proposition and a Keynesian alternative'. *Economic Journal*, 98:498–503

Sawyer, K. R. 1989. 'Non-parametric estimation and testing of econometric models of duration: an application to the Australian Longitudinal Survey'. *Australian Journal of Statistics*, 31A:143–64

Sawyer, K. R. and M. Alauddin. 1989. 'A microeconometric search for a Phillips curve using industry wage data'. University of Melbourne Department of Economics. Working Paper 170

Sims, C. A. 1972. 'Money, income and causality'. *American Economic Review*, 62:540–52

Sims, C. A. 1980. 'Macroeconomics and reality'. *Econometrica*, 48:1–47

Stock, J. and M. Watson. 1988. 'Variable trends in economic time series'. *Journal of Economic Perspectives*, 2:147–74

Thursby, J. G. 1981. 'A test strategy for discriminating between autocorrelation and misspecification in regression analysis'. *Review of Economics and Statistics*, 63:117–23

Tobin, J. 1972. 'Inflation and unemployment'. *American Economic Review*, 62:1–18

Trzincka, C. 1986. 'On the number of factors in the arbitrage pricing model'. *Journal of Finance*, 51:347–68

Tsurumi, H and Y. Tsurumi. 1983. 'US–Japan automobile trade: a Bayesian test of a product life cycle'. *Journal of Econometrics*, 23:193–210

Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley

Ullah, A. (ed.). 1989. *Semiparametric and Nonparametric Econometrics. Studies in Empirical Economics*. Physica; Springer

Wulwick, N. J. 1987. 'The Phillips curve: which? Where? To do what? How?' *Southern Economic Journal*, 53:834–57