

# The experiment in macroeconometrics

John Aldrich and Anna Staszewska

---

**Abstract** This paper examines the *experiment* in macroeconometrics, the different forms it has taken and the rules that have been proposed for its proper conduct. Here an ‘experiment’ means putting a question to a model and getting an answer. Different types of experiment are distinguished and the justification that can be provided for a particular choice of experiment is discussed. Three types of macroeconomic modelling are considered: the Cowles (system of equations) approach, the vector autoregressive model approach and the computational experiment.

**Keywords:** experiment, impulse response analysis, *ceteris paribus*, structural invariance

**JEL Classifications:** B41, C5, E37

## 1 INTRODUCTION

This paper examines the *experiment*, the forms it has taken and the rules which have been proposed for its conduct, in three varieties of macroeconometrics, based on the computational experiment, the unstructured vector autoregressive model and the system of equations (Cowles). The experiments are not material experiments involving the manipulation of the material world but what Morgan (2002: 42–3) calls ‘model experiments’ and describes as ‘glorified’ thought experiments, glorified because they are too complicated to be done in the head. In such experiments a question is put to a model and reasoning and calculation are involved in getting an answer. It is sometimes said that in such experiments anything goes: ‘Any thought experiment is permissible’, writes Samuelson (1976: 521), though he adds, ‘some thought experiments seem less interesting than others’. Among the issues we consider are whether all experiments are permissible and what makes a permissible experiment interesting.

Model experiments are used in the research tradition associated with Kydland and Prescott but in recent years these authors have given the ‘experiment’ a wholly new prominence – especially in their reflective essays (1991 and 1996) – so that the ‘computational experiment’ has come to stand

for their distinctive style of macroeconometrics much as the ‘econometric model’ stood for an earlier style. However, experiments are to be found in other varieties of macroeconometrics, if not so well billed. Our first task is to bring out the experiment in the different styles and devise categories that will apply to all; thus the suggestive distinction between ‘using theory’ and ‘developing theory’, which appears to be specific to the literature on the computational experiment, will be applied more generally. Another task will be to identify each approach’s code of practice, the ‘local’ ‘within-approach’ rules on experiments. Such a code is implicit when Koop *et al.* (1996) justify their new form of impulse response analysis for the vector autoregressive approach against established forms. Noting that the impulse response function is the ‘outcome of a conceptual experiment’, they argue (1996: 120) – echoing Samuelson – that in choosing between experiments, ‘The point ... is not whether a particular choice is right or wrong, but whether the choice is economically meaningful or interesting.’ In section 4 below their notion of ‘meaningfulness’ is examined. Naturally experimentalists don’t only weigh the best form of experiment for their own approach, they also judge experiments conducted according to other codes – thus Cooley and LeRoy (1985) are led to criticize VAR experiments from the standpoint of the system of equations approach.

The approaches we consider, based on the computational experiment, the unstructured VAR and the systems of equations, are ideal types but they have been important in macroeconomics and the elements that make them up can be recombined to generate other forms of macroeconometrics which have been practised and more that might be conceived: the econometrics of Lucas and Sargent (1981), like the computational experiment, is based on new-classical macroeconomic theory but uses estimation to obtain parameter values instead of calibration; the structural VARs of Bernanke (1986) and Blanchard and Quah (1989) combine features of the system of equations approach and the VAR approach. Incidentally we refer to varieties of *macroeconometrics* out of courtesy to their macroeconomist creators who say they are doing econometrics and regard each other as competitors. The appropriateness of the label is not a pressing issue here; see Hoover (1995) for some interesting observations on the matter.

The ‘logic’ of question and answer – the meaningful and the interesting – as it has developed in the three varieties of macroeconometrics has a hinterland. Although the writing we engage with was produced in the main by ‘working’ econometricians and economists, there is related work in the methodology of economics on the use of *ceteris paribus* conditions; see e.g. Whitaker (1991) and Hausman (1992), and in the philosophy of science on ‘mechanisms’, see e.g. Machamer *et al.* (2000) and Woodward (2002). Much of the literature on models is about how they are used in science, how they ‘function in the large’. We do not consider this question but ask, *given* that models – quite specific ones – *are* used, what counts as an effective use?

Models and experiments are often studied in juxtaposition – see e.g. Mäki (2005) and Morgan (2005) – but the experiments are usually ‘lab experiments’ and, while there are points of contact with the present paper, the basic concerns are quite different. Put another way, the subject matter of the present paper can be located in the boxes contained in Table 1 of Boumans and Morgan (2001) and interpreted as a microscopic study of what is in those boxes but the table’s authors are more interested in what distinguishes those boxes from other boxes! There is a more relevant study of the ‘experimental’ in Boumans (2003) – indeed there is some overlap in the discussion of the system of equations approach – Haavelmo and Cowles.

## 2 QUESTIONS AND ANSWERS

We start with some general remarks on questions and answers, some terminology and some preliminary distinctions. Following Kydland and Prescott (1996: 75), ‘running an experiment’ will mean using a model to answer a question; macroeconometricians of other persuasions do-not-use the word ‘experiment’ so freely and so will not immediately recognize how many of their activities fall under this broad heading. We adopt this perspective as an ‘experiment’ in sense 1 (the oldest) in the *Oxford English Dictionary*, viz. ‘the action of trying anything, or putting it to proof; a test, trial’. It is worth trying because it enables us to establish common ground between the deviant usage of the computational experimentalists and ‘ordinary’ language.

In the most basic situation the question comes from curiosity about some aspect of the world to which the model refers; the model is applied to produce an answer which will satisfy our curiosity. In macroeconometrics the answering of a ‘what happens next’ question is called forecasting and the answering of a ‘what happens when’ question is called simulation; these particular application experiments may be called descriptive experiments. If the model embodies some theory then we can say with Kydland and Prescott (1996: 71) that we are ‘using theory’. The econometric models considered below are constructed using data and have a specific reference to the universe that produced the data, although their creators may make claims about a wider, less well defined universe. Forecasting is more tied to the ‘here and now’ than simulation, yet there is little point in asking what will happen in some improbable set of circumstances; the unrealism of a question will be held against it. The models found in all three versions of macroeconometrics are competent to answer such questions and this may be considered their primary function. The details vary according to the approach (see sections 3–6 below). The emphasis given to the two activities is not uniform – the computational experiment is more associated with simulation than with forecasting – and then again models differ enormously in the range of questions they can answer; the computational experiment is often tailored to a single question, while systems of equations are usually

general purpose models that can answer many questions. These differences appear to be accidental rather than inherent in the approaches.

The primitive function of finding an answer to a specific question about the world 'outside' the model generates further functions. Once a model is recognized as an answer-generating machine which may be good for more than a single use, we want to know what questions it can answer and how it answers certain standard questions; the answers to these questions constitute the 'properties' of the model. The properties can describe what usually happens but they do not have to, for they can describe what rarely or even what never happens; some application experiments are not at all descriptive. Models are used for understanding the process being modelled, not just as oracles. Analytical experiments that investigate how the model works are instructive in so far as they show something about the working of the model. We will see that the questions best designed to elucidate the working of the model may be unrealistic and far-fetched.

In answering these questions we apply a model in the belief that it is relevant and obtain an answer in which, consequently, we have some confidence. There is no dissent about what is needed to make an answer credible; it should be a valid derivation from a credible model. Kydland and Prescott's (1996: 83) statement, 'If the theory is strong and the measurements good, we have confidence that the answer for the model economy will be essentially the same as for the actual economy' interpreted generously would command assent from other users of economic theory, while 'good measurements' would satisfy the structureless VAR modeller. Disagreements appear when content is given to 'strong' theory and 'good' measurements. Another principle about application experiments that would command assent is expressed by Kydland and Prescott (1996: 73) in the form, 'The selection and construction of a particular model economy should not depend on the answer provided'. If the model is chosen because it delivers a particular answer and has no independent credibility then the answer's credibility is not increased by being derived from such a model. However, while selecting a model on the basis of the particular answer it provides is not sensible, selecting a model on the basis of the kind of question it can answer is obligatory, for only models competent to answer the question are of interest.

There are many reasons for running experiments and the identical experiment may be run for different reasons – the 'so what?' question prompted by any question–answer pair can usually be answered in several ways. Besides experiments in which the model is applied to discover something directly about the world, there are experiments whose point is to challenge the model. Kydland and Prescott (1996) write about 'developing' theory when the interest is in the theory and the question and answer scheme is a tool for investigating the theory. Thus as well as putting questions in good faith – out of ignorance – we may ask questions, *knowing* the answer,

in order to evaluate the model's ability to produce that answer. If the model produces the correct answer in such a test experiment, our confidence in its ability to answer genuine questions increases. If models were non-stochastic, testing and estimation could simply proceed by question and answer: does the answer produced by the model match the answer found outside; does this parameter value produce the right answer? In stochastic models the answer takes the form of a probability distribution and the matching of the model outcome with the answer found 'out there' is more complex. Test experiments are not the only model-challenging experiments. Kydland and Prescott (1996: 70) mention experiments 'concerned with the sensitivity of previous findings to the introduction of some features of reality from which previous studies have abstracted'. Such experiments are routine in all three approaches. However, test experiments are the most discussed of the experiments charged with developing theory; they are examined in section 6 below.

Although the entire paper is an elaboration of these points, it may do no harm to give an example which will illustrate some of them and indicate how disputes about experimental reasoning may arise. 'If kangaroos had no tails, they would topple over' (borrowed from Lewis (1973: 1)) is the headline finding of an analytical experiment in mechanics designed to show how the kangaroo works as a mechanical system. The sound observation that the experiment is not realistic – under what circumstances would kangaroos be deprived of their tails – is not a telling objection, for such realism is not required for the argument to be an effective demonstration. Change now the universe of discourse to evolutionary biology and contemplate a race of kangaroos without tails, animals identical to kangaroos except they do not have tails. As the mechanical principles still obtain, such creatures would be unviable and, not existing, they would not fall over – or keep their balance. One may want to rule impermissible any experiment involving tail-less kangaroos. Or one might contemplate a race of near-kangaroos without tails. To be viable these creatures would have had to have solved the stability problem in some way and so such tail-less near-kangaroos would not topple over. Here is a case of an unrealistic experiment which seems worth contemplating. However, depending on how precisely the experiment is specified, it may be permissible or impermissible; if it is permissible the outcome as reported may be correct or it may be incorrect.

We now apply these ideas to the three approaches to macroeconometrics, beginning with the oldest, the system of equations approach. This produced the richest literature on application experiments, including a discussion of what is permissible and accounts of different analytical and descriptive experiments. It was the starting point for the later approaches in two distinct and actually opposed ways: its perceived failure as a practical project spurred their development, while that development fed off the conceptual capital it built up.

### 3 INDIVIDUAL-EXPERIMENTS AND MARKET-EXPERIMENTS

Tinbergen pioneered the systems of equations approach in macroeconometrics in the 1930s. The statistical foundations were overhauled by Haavelmo (1944) and a Cowles monograph – Koopmans (1950) – developed the statistical apparatus, while another – Klein (1950) – revived the Tinbergen project with the new statistical tools and Keynesian economic theory; see Morgan (1990) for a full account. The modelling activity involved two stages: the construction of a credible econometric model; the investigation of its properties and its application to forecasting and policy simulation. This division of labour found classical expression in the diptych Klein and Goldberger (1955) and Goldberger (1959). In this section we consider the experiments performed on the constructed model, deferring discussion of the model-challenging experiments which belong to the construction stage to section 6. Note that we do not consider it necessary to distinguish between Haavelmo and the Cowles writers, as Boumans (2003: 314) does, and that by homogenizing the econometricians and the theorist Patinkin we merge some of the boxes in the Boumans and Morgan table.

Tinbergen's equations reproduced the standard experiment of multiple regression; in the table of Boumans and Morgan (2001) this appears as 'indirect control (statistical)'. Yule (1899: 251) described how a causal 'factor' is allowed to vary subject to the condition that the other factors are 'kept constant'. This remained the standard experiment when R.A. Fisher (1925) changed the statistical model in which the experiment was analysed; for this history see Aldrich (1995 and 2005). This analytical experiment shows the separate influence of the factors; several such experiments can be used to find the response to a realistic combination of factors. There is another realistic question: how does the dependent variable change as one of the factors changes, recognizing that the other factors will usually change as well? Geary (1963) argued that this is the practically important question, not the analytical question. If Geary's descriptive experiment is tackled from the standpoint of the separate analytical experiments, the co-movements of the factors must be specified. However, the descriptive experiment is easily run by performing a simple regression of the dependent variable on the factor. The disagreement indicates the range of interests in regressions: Geary was interested in forecasting, Yule in the effects of a change in policy and Fisher (1925: 130) in the science of how rainfall varies with latitude and longitude. See Aldrich (1997).

The system of equations approach to macroeconometrics and the system of equations approach to macroeconomic theory evolved in parallel and the equations of the econometric model mirror the equations of economic theory. The common structure is described in Patinkin's *Money, Interest and Prices* (1956) under the headings of 'individual-experiments and

market-experiments' (p. 15) and 'derivatives in economic analysis' (pp. 275–88). The economic theorist was describing the pattern the econometricians – especially Haavelmo – had abstracted from the practice of economic theorists. For his illustrations Patinkin draws on general equilibrium theory, partial equilibrium analysis and simple macroeconomic analysis. The experiments are analytical experiments designed to show the working of the model under investigation. Before describing the Haavelmo–Patinkin scheme we make a small digression.

Haavelmo (1944) – 'The probability approach in econometrics' – finishes by extending statistical theory to cover the situations economists study but begins by analysing the situation of the economist as a student of the economy. This discussion has a lot about experiments and most of it will be of no concern to us. But to avoid misunderstanding it may be best to mention two prominent but – for our purposes – irrelevant themes. The starting point for the enterprise is the contrast, going back at least to Mill, between the active experimentation practised in many sciences and the 'passive observation' practised in economics; this side of Haavelmo's work is emphasized by Boumans and Morgan (2001) and Boumans (2003). Haavelmo was aware of Fisher's famous book *The Design of Experiments* (1935) which treated experiments in which 'things' are manipulated from a statistical viewpoint. Given this background, it is rather surprising that the phrase 'design of experiments' applied to *economics* appears so often, or even at all, in the 'Probability approach'. Haavelmo (1944: 14) uses the expression in an entirely idiosyncratic way to refer to the operational definition of model entities, the rules of correspondence between them and 'real economic phenomena'. The experiments we are considering are not experiments as physical manipulation or experiments as measurement, but experiments as manipulation of model entities; Morgan (2002) calls them 'experiments *within* models' or 'model-experiments' which makes the present essay an investigation into the 'design of experiments' – model-experiments.

To return to the Haavelmo–Patinkin experimental scheme, consider a model of an isolated market as in Haavelmo (1944: 99ff) or Patinkin (1956: 275ff):

$$q_t^d = \gamma_0^d + \beta^d p_t + \gamma^d i_t + \varepsilon_t^d$$

$$q_t^s = \gamma_0^s + \beta^s p_t + \gamma^s w_t + \varepsilon_t^s$$

$$\varepsilon_t \sim IN(0, \Sigma)$$

$$q_t^s = q_t^d = q_t$$

(This Marshallian model is a paradigm for methodologists too – see e.g. Hausman (1992: 48ff) and Morgan (2002: 44ff.) The demand and supply equations are specified in terms of (unobserved) quantity demanded,  $q_t^d$ , and quantity supplied,  $q_t^s$  where  $p_t$  is price,  $i_t$  is consumers' income affecting

demand and  $w_t$  is a weather variable affecting supply;  $\varepsilon_t^d$  and  $\varepsilon_t^s$  are jointly normal ‘random shifts’ affecting demand and supply. The last equation is a ‘market relation’ equating quantities demanded and supplied to the observed quantity sold,  $q_t$ . The income and weather variables are ‘assumed to remain fixed in repeated samples’ (1944: 99). The ‘economic meaning’ of this is that each is an ‘autonomous parameter fixed by forces external to the economic sector under consideration’ (p.86). The transformation of ‘autonomous parameter’ into ‘exogenous variable’ is described in Aldrich (1993).

The demand and supply equations describe the results of what Patinkin calls *individual-experiments* where the purpose of the analysis is ‘to conduct conceptual experiments with the individual – or group of individuals – in which arbitrary changes in one or more independent variables are made and the effect of these changes on the dependent variables observed’. In the case of the consumer, ‘What we take as given and not subject to explanation – the “independent variables of the analysis” – are the individual’s tastes and initial endowments, the prices with which he is confronted, and his desire to transform his initial endowment into an optimum one.’ As an example, Patinkin takes the excess demand curve which shows ‘how the amount of excess demand varies with changes in the price of  $X$  – other independent variables being assumed constant’.

In *market-experiments* the unit of investigation is not the individual or group of individuals but ‘the organized market operating under conditions of perfect competition’. The aim of analysis at this level of investigation is to investigate the effects of changes in the independent variables on the equilibrium price of the market. In terms of the example the ‘market experiments’ are expressed in the relationships between the ‘dependent’ variables  $q_t$  and  $p_t$  and the ‘independent’ variables,  $i_t$  and  $w_t$  and the random shifts,  $\varepsilon_t^d$  and  $\varepsilon_t^s$ . These reduced form equations are obtained by solving the structural form equations for  $q_t$  and  $p_t$

$$\begin{aligned} q_t &= \frac{\beta^d (\gamma_0^s + \gamma^s w_t + \varepsilon_t^s) - \beta^s (\gamma_0^d + \gamma^d i_t + \varepsilon_t^d)}{\beta^d - \beta^s} \\ &= \pi_0^q + \pi_w^q w_t + \pi_i^q i_t + v_t^q; \\ p_t &= \frac{(\gamma_0^s + \gamma^s w_t + \varepsilon_t^s) - (\gamma_0^d + \gamma^d i_t + \varepsilon_t^d)}{\beta^d - \beta^s} \\ &= \pi_0^p + \pi_w^p w_t + \pi_i^p i_t + v_t^p \end{aligned}$$

The coefficients,  $\pi_w^q$ ,  $\pi_w^p$ , etc., describe the reaction of quantity and price to changes in the weather and in income. The reduced form also supports descriptive experiments, e.g. predicting the value of  $p_t$  associated with a realistic combination of  $w_t$  and  $i_t$ .

The equations can support other experiments: e.g. the effect of a change in the value of  $\gamma^d$  on the distribution of  $q_t$  and  $p_t$  can be investigated using the information that  $\pi_i^q = \beta^d \gamma^d / (\beta^d - \beta^s)$  and  $\pi_i^p = \gamma^d / (\beta^d - \beta^s)$ . Such questions cannot be answered knowing only the values of the  $\pi$ 's and not how they are formed. This insight was central to the Cowles case against a purely statistical approach concerned only with estimating the  $\pi$ 's; see e.g. Marschak (1950). The Cowles econometricians had a larger repertoire of experiments than Patinkin; there were realistic experiments as well as analytical experiments on the existing structure but there were also analytical and realistic experiments based on a possibly reconstructed structural form. Marschak (1949) considered taking an estimated supply and demand model and asking what happens if market clearing is replaced by a system of price control – replacing the condition  $q_t^s = q_t^d = q_t$  by  $q_t = \min(q_t^s, q_t^d)$ . The change implies a new joint distribution for  $q_t$  and  $p_t$ , superseding the original reduced form.

Patinkin was extending the analysis of Chapter 1 of Samuelson (1947) which had considered only market experiments; Patinkin's purpose in distinguishing between market and individual experiments was to show how it is possible to describe the response of demand to price while maintaining that, 'it is meaningless to inquire as to the effects of an arbitrary change of price, for, by assumption, such a change can now only occur as the result of a prior change in one of the independent variables'. Whitaker (1991: 396) cried impermissibility, finding a 'logical inconsistency' in Patinkin's analysis of individual experiments: for him *ceteris paribus* is used as an instruction 'to hold hypothetically constant some members of a set of *necessarily* covarying variables while changes in the others are contemplated'. Whitaker considers *ceteris paribus* a fudge and confines individual experiments to a 'heuristic role in aiding comprehension of the attainment and character of general equilibrium, even though they are not part of the theory's logical structure'.

Neither Whitaker nor Patinkin described what some econometricians and economic theorists have considered a logical inconsistency in experimentation: viz., acting on the false assumption that, when one equation of a system is changed, the others will remain unchanged. In the kangaroo fable there was the question of how much of the kangaroo could be meaningfully changed. Frisch (1938) and Haavelmo (1944) were interested in the status of equations fitted to data and how one parametrization came to have a privileged status over other observationally equivalent representations. Demand and supply relations are 'autonomous' in the Frisch–Haavelmo sense that one equation could be altered without necessarily altering the other equation – e.g. the demand equation could change because the circumstances of consumers change while the supply equation remains the same; by contrast, the parameters of each reduced form equation depends on the parameters of both demand and supply equations. The idea of autonomy was part of the econometrician's notion of a 'structural

equation': structural equations 'can, at least in theory, be changed one by one, independently' (Koopmans *et al.* 1950). The concept of autonomy is discussed by Aldrich (1989), Cartwright (1989) and Boumans (2003); as 'modularity', it has been reintroduced by Woodward (2002).

Frisch and Haavelmo were interested in relations 'found' in the data but in an era when Keynesian theory was being developed theorists were interested in the status of relations proposed by other theorists; Schumpeter (1954: 1182, fn. 25) praised Frisch's discussion of autonomy as a contribution to 'the pure logic of modern theory'. Keynes (1936: 179–81) argued that the classical theory of interest is a 'nonsense theory' because the demand curve for capital and the supply curve are constructed on the assumption that income is constant but that if either curve shifts, 'then, in general, income will change; with the result that the whole schematism based on the assumption of a given income breaks down'. Keynes's argument is based on the theory that the basic relationships are the marginal efficiency of capital, the propensity to consume and liquidity preference and that, while the classical supply and demand curves can be derived from them, they cannot be derived in a way that produces a pair of autonomous relations. The IS-LM construction also reduced the system to two relationships but the derived relationships are designed to be autonomous because they are constructed from non-overlapping sets of basic relationships. Against Keynes and as part of his 'methodological criticism of the neo-Cambridge school' Leontief (1937: 350) argued on the basis of general equilibrium theory that Keynes's aggregate supply and demand equations are not autonomous:

[Mr. Keynes] obviously implies a situation where one of the two functions 'shifts' while the other retains its shape and position unchanged; a situation which is, because of the fundamental interdependence of the two functions, highly improbable if not to say impossible. The methodological danger of potential misinterpretation, which is inherent in any implicit statement, here actually leads to an open logical inconsistency.

It will depend on the precise form of the model whether the experiments Keynes and Leontief criticize are impossible – Keynes (pp. 179–80) mentions the possibility that the position could be 'saved' by 'some complicated assumption'. However when the misconceived experiment is analysed in terms of the true specification the experiment is revealed as not what it seems and to involve at best some very contorted changes.

The issue of autonomous relations re-emerged in a later round of macroeconomic controversy in Sargent's 'The observational equivalence of natural and unnatural rate theories of macroeconomics' (1976) and Lucas's 'Econometric policy evaluation: a Critique' (1976). The latter, a criticism of the application experiments carried out on Keynesian dynamic simultaneous equations models (see below), is a variation on the theme of presumed

autonomy; adapting Keynes’s words, ‘the whole schematism based on the assumption of a given *expectation* breaks down’. The Lucas critique became an item in the literature in a way in which the Keynes and Leontief arguments did not. Whether the inconsistency Lucas identified matters for policy evaluation has been debated by Sims (1982) and whether the theoretical possibility is reflected in any data has been considered by Ericsson and Irons (1994). Kydland and Prescott (1991) take the Lucas critique for a fact and indeed the ‘general recognition that policy-invariant behavioral equations are inconsistent with the maximization postulate in dynamic settings’ is one of the reasons they give for the decline of the systems of equations approach.

The simplest case of non-autonomy seems not to have come up in controversy – perhaps it is too transparently silly. The variables subject to ‘arbitrary changes’ in the individual demand experiment are  $p_t$  and  $i_t$  but the equation could equally be written in terms of variables constructed from them, e.g.  $r_t = p_t + i_t$  and  $s_t = p_t - i_t$ :

$$\begin{aligned} q_t^d &= \gamma_0^d + \beta^d p_t + \gamma^d i_t + \varepsilon_t^d \\ &= \gamma_0^d + \left(\frac{\beta^d + \gamma^d}{2}\right)(p_t + i_t) + \left(\frac{\beta^d - \gamma^d}{2}\right)(p_t - i_t) + \varepsilon_t^d \end{aligned}$$

Holding one of each pair of variables, say  $r_t$ , fixed while changing the other,  $s_t$ , involves changing *both* variables of the other pair,  $p_t$ ,  $i_t$ . Analytical experiments would be based on the original pair of variables which have an economic interpretation or identity because they appear in the reasoning on which the experiments are based.

In practice fitted macro-models are invariably dynamic and so we need to notice what this extra dimension brings to the static models we have been considering. The structural form of the linear dynamic simultaneous equations model in which the action of the predetermined (exogenous  $z_t$  and lagged endogenous  $y_{t-1}$ ) variables and shocks  $\varepsilon_t$  on the endogenous variables  $y_t$  is given by

$$B_0 y_t = B_1 y_{t-1} + \Gamma z_t + \varepsilon_t, \quad \varepsilon_t \sim IN(0, \Sigma)$$

where  $y_0$  is fixed.

(In practice the dynamics would be richer than first- order.) The simplest individual experiments involve reading the coefficients of the equations, the elements of the matrices,  $B_0$ ,  $B_1$  and  $\Gamma$ . Market experiments are based on the reduced form or on the final form, depending on how far ahead the experimenter wishes to look. The *reduced form* is, say

$$\begin{aligned} y_t &= B_0^{-1} B_1 y_{t-1} + B_0^{-1} \Gamma z_t + B_0^{-1} \varepsilon_t \\ &= \Pi_1 y_{t-1} + \Pi_0 z_t + v_t \end{aligned}$$

The matrix  $\Pi_1$  describes the influence of last period's state  $y_{t-1}$  on this period's state. The matrix  $\Pi_0$  describes the impact of the exogenous variables on the endogenous variables; the term  $B_0^{-1}$  describes the impact of the structural shocks on the endogenous variables. In the *final form* the current value  $y_t$  is expressed in terms of current and past values of the exogenous variables and the shocks:

$$y_t = \Pi_1^t y_0 + \sum_{s=0}^{t-1} D_s z_{t-s} + \sum_{s=0}^{t-1} C_s \varepsilon_{t-s}$$

where  $D_s = \Pi_1^s \Pi_0$  and  $C_s = \Pi_1^s B_0^{-1}$ . The matrix  $D_s$  describes the influence of past exogenous variables  $z_{t-s}$  on this period's state,  $y_t$  and an element gives the effect of a unit change in one of the exogenous variables on one of the endogenous variables  $s$  periods later holding all the other variables fixed both in the original period and in all the intervening ones.

Goldberger's (1959: 2) object was to calculate the 'total, indirect relationships' in the final form given that the 'partial, structural, direct relationships' in the structural form of the Klein–Goldberger (1955) model are already known. The final form supports both descriptive experiments and analytic experiments; calculating the effects of policy packages is an example of the former and the interpretation of the dynamic multipliers, the  $d_s$ , as representing the effect of a unit change in a variable holding the others fixed, is an example of the latter. In the Cowles approach little attention was paid to experiments involving the random shocks, although the market model has identifiable supply shocks and demand shocks. However, the final form could be used to follow out the effects of a shock to one of the elements of  $\varepsilon_t$ , using the forms  $C_s = (B_0^{-1} B_1)^s B_0^{-1}$ . However, it would make no sense to consider the effect of a shock to one of the elements of the reduced form error  $v_t$  holding the other elements fixed because a change to one of the elements of  $v_t = B_0^{-1} \varepsilon_t$  only comes about because of a change to some element(s) of  $\varepsilon_t$  but this will affect other components of  $v_t$  as well. For this reason the experiments conducted by VAR modellers (see section 4 below) appear to be an academic exercise. Adelman and Adelman (1959) considered the dynamic properties of the Klein–Goldberger model by subjecting it to 'realistic' random shocks, drawing the vectors  $\{\varepsilon_t\}$  from the distribution estimated by Klein and Goldberger. The Adelmans were interested in appraising the model as a representation of the US economy in a way additional to work on the individual equations. The Adelmans described the properties of the model by calculating the descriptive statistics used by the National Bureau of Economic Research for business cycles. They concluded, 'it would appear that the shocked Klein–Goldberger model approximates the behavior of the United States economy rather well' (Adelman and Adelman 1959: 620). The realistic applications experiment could be read as a test experiment; we return to test experiments in section 6 below.

The Cowles project combined economic theory and statistical inference. The VAR approach retained statistical inference but discarded economic theory; the computational experiment retained economic theory and discarded statistical inference. In the next two sections we review the application experiments associated with these approaches and return to the Cowles project in section 6 when we describe model-challenging experiments.

#### 4 VARS: 'THE BEST DESCRIPTIVE DEVICE'

In 'Macroeconomics and reality' (1980) Sims argued that the system of equations approach had failed because in practice it was not possible to recover the individual experiments believed to inform the data: 'claims for identification in these models cannot be taken seriously' (p. 1). Sims rejected the project of formulating individual experiments and, with it, the matching of equations in the model to be estimated with an economic theory model, and chose instead to 'estimate large-scale macromodels as unrestricted reduced forms, treating all variables as endogenous' (p. 15). Naturally the range of experiments narrowed to match; gone are the reconstruction experiments of the kind envisaged by Marschak or Lucas. On the other hand, Sims argued that routine policy analysis did not involve that kind of reconstruction.

For his unrestricted reduced forms Sims chose the vector autoregressive model:

$$x_t = \Phi x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IN(0, \Omega)$$

where  $x_0$  is fixed (for simplicity again we consider the first order process). Here  $x_t$  is the vector of variables and  $\Phi$  the coefficient matrix. The variance matrix  $\Omega$  registers the contemporaneous covariation of the random shocks contained in the vector  $\varepsilon_t$ . In macroeconomic applications the off-diagonal elements of  $\Omega$  are typically non-zero.

The standard objection from the Cowles direction – see e.g. Cooley and LeRoy (1985: 304–6) – is that non-structural models like this one which only 'capture the probabilistic characteristics of the data under examination' cannot answer questions about the effects of interventions: the only permissible experiment is the forecast, given the present position  $x_t$  what is the distribution of  $x_{t+1}$ ? VAR modellers, however, have wanted more than an oracle; they have wanted to understand the mechanism of the economy.

In the system of equations approach the standard analytical experiment was not at all realistic. The goal of experimental design for the VAR has been the experiment which is both analytical and realistic yet relies on a minimum of structural information. Sims (1980: 20–1) judged the available experiments unilluminating: 'It is especially difficult to make sense of [autoregressive systems] by examining the coefficients in the equations themselves.' Information about the behaviour of the system can be extracted from functions of these coefficients such as the eigenvalues. However, Sims

argues, ‘The best descriptive device appears to be analysis of the system’s response to typical random shocks.’ Solving the VAR in terms of the shocks and the initial conditions

$$x_t = \sum_{i=0}^{t-1} \Phi^i \varepsilon_{t-i} + \Phi^t x_0$$

gives a representation corresponding to the final form of the dynamic simultaneous equations model of section 3 above. Following the example of the analytical experiments of section 3 we would set  $\varepsilon_1 = (0, 1, \dots, 0)'$  and  $\varepsilon_s = (0, 0, \dots, 0)'$  for  $s > 1$  and trace over time the way a random shock in the second equation at time 1 works through the system. Koop *et al.* (1996) call this the *traditional approach* although there is no great tradition of its use in VAR analysis.

For Sims the traditional experiments are unrealistic; there is nothing ‘typical’ about these shocks. The vector  $\varepsilon_1 = (0, 1, \dots, 0)'$  does not look like a drawing from the shock distribution with non-diagonal  $\Omega$ . Koop *et al.* (1996: 128) are quite explicit: holding the other shocks fixed ‘is not likely to be economically relevant or meaningful; not the sort of shock that the Data Generation Process will often generate’. Sims (1980: 22) reports and recommends:

The residuals are correlated across equations. In order to be able to see the distinct patterns of movement the system may display it is therefore useful to transform them to orthogonal form. There is no unique best way to do this. What I have done is to triangularise the system, with variables ordered as  $M, Y, U, W, P, PM$ .

The triangularization process creates orthogonal errors  $v_t$  because associated with  $\Omega$  there is a triangular matrix  $T$  such that  $T\Omega T' = I$  and so

$$Tx_t = T\Phi x_{t-1} + T\varepsilon_t, \quad E(T\varepsilon_t \varepsilon_t' T') = T\Omega T' = I$$

$$\text{or, } x_t = \sum_{i=0}^{t-1} \Psi_i v_{t-i} + \Phi^t x_0$$

where  $v_t = T\varepsilon_t$  and  $\Psi_i = \Phi^i T^{-1}$ . With the orthogonal errors,  $v_t$ , experiments of the  $(0, 1, \dots, 0)$  variety were secure: the errors did not move together and each component of  $v_t$  could still be associated with a variable, a component of  $x_t$ .

The transformed system with a triangular coefficient matrix on the vector  $x_t$  has a place in the family of simultaneous equations models as the recursive model and it can be interpreted as a structural model with  $v_t$  the structural error. Sims recognized that other orderings were possible but thought the ordering he had chosen was reasonable. Koop *et al.* (1996: 128) objected to Sims’s procedure:

In the absence of other a priori information (e.g., a logical causal ordering among the variables in the VAR) there is an arbitrariness implicit in such a transformation which can lead to difficulties in interpretation of the impulse response functions.

The difficulties are presumably that different orderings ascribe different responses to a shock to the same component of  $x_t$ . Sims (1981) examined the sensitivity of the conclusions to the choice of ordering; they could be very sensitive.

To obtain an experimental procedure free of arbitrariness Koop *et al.* propose *generalized impulse response analysis*. This assumes that the other shocks, i.e. those to other components of  $\varepsilon_t$ , take values in accordance with the contemporaneous correlations implied by the matrix  $\Omega$ ; it appears to be the method Sims rejected on the grounds that 'it is difficult to see the distinct patterns of movement the system may display'.

Reviewing the three experiments, the traditional, the orthogonalized and the generalized, all try to illuminate the way the system works but they come with different degrees of realism, clarity and arbitrariness. It is interesting to compare these experiments with those described in section 3 for regression and the simultaneous equations model; the traditional and the generalized correspond to the Yule–Fisher and Geary experiments and a parallel to the Sims procedure exists in the form of regression with orthogonalized regressors. Koop *et al.* identify the 'economically meaningful or interesting' with that which is likely to happen while the analytical experiments of section 3 are meaningful because they involve variables with an economic identity and interesting because they show how the system works. In the VAR case the elements of  $\varepsilon_t$  have a meaning derived from their position in the VAR specification; in some cases the element may be given an interpretation (identified) as a surprise to the corresponding variable. However, if the VAR is interpreted as a reduced form, then the component has no straightforward economic identity. In showing how the system works the regression model treats the transmission of influence from the exogenous to the endogenous variables; the mode of the generation of the exogenous variables is outside the system. Neither Sims nor Koop *et al.* treat the VAR as a model recording only the way influence is transmitted from  $\varepsilon_t$  to  $x_t$  for  $\Omega$  is regarded as much a part of 'the system' as  $\Phi$ . In terms of the discussion of section 3 the components of  $\varepsilon_t$  are not treated as autonomous. In the structureless VAR set-up where there is no *a priori* information the generalized impulse response analysis has the virtue that nothing is contemplated in the experiment that has not been experienced in the sample. There is however the constitutional problem in this set-up that it is not known whether the shock whose consequences are being traced is a natural shock or an artificial one – in other words, what the economic meaning of the experiment is.

The structural imperative demands an economic identity for every shock. The structural VAR models, as presented e.g. by Blanchard and Quah (1989) and Christiano *et al.* (1999), reverse the order found in Sims (1980): the impulse responses carry the economic theory and the autoregressive model provides a computationally convenient way of expressing them. The question, ‘What have we learned about monetary policy shocks?’ has replaced the Cowles question, ‘What have we learned about the demand for money?’. The challenge of the ‘shocking story’ for statistical inference is described in Levchenkova *et al.* (1998) and Staszewska (2004).

## 5 ‘USING THEORY’ AND THE COMPUTATIONAL EXPERIMENT

‘Any economic computational experiment’ write Kydland and Prescott (1996: 70) involves the following steps: ‘pose a question; construct a model economy; calibrate the model economy; and run the experiment’. Studies divide into those ‘using theory’ and those ‘developing theory’ (see above section 2) but the work involved in each is largely the same. In the first category come policy evaluation issues, such as the welfare consequences of the 1986 US tax reform. In the second come experiments designed to produce conclusions about the model economy, including grounds for changing it; those experiments are considered in the next section.

The ‘theory’ looks different from that which figures in the system of equations approach; it is ‘Jevonian’ with a single protean representative agent rather than ‘Marshallian’ where each equation describes the behaviour of a set of agents. Further, although the model economy is constructed from the same modules – ‘preferences’ and ‘technology’ – as underlie Patinkin’s individual experiments, the technique is quite different because intertemporal optimization in even quite simple settings generates no ‘behavioral equations’, no objects comparable to the vehicles for Patinkin’s individual-experiments.

The detailed specification of the model depends on the question posed but the following example of the welfare implications of a change in the size of technology shocks is typical. The model economy is a dynamic stochastic general equilibrium model consisting of identical agents facing uncertainty about future technology and solving a dynamic consumption/labour supply problem. The agent’s problem is to

$$\max E \left[ \sum_{t=1}^{\infty} \beta^{t-1} U(c_t, 1-h_t) \right]$$

subject to  $c_t + i_t = z_t f(k_t, h_t)$

$$z_{t+1} = z_t^\rho \varepsilon_{t+1} : 0 < \rho < 1 : \ln \varepsilon_t \sim IN(0, \sigma^2)$$

$$k_{t+1} = k_t(1 - \delta) + i_t : k_1 \text{ given}$$

(This particular formulation is from Hartley *et al.*'s useful *Reader* (1998: 44).) The technological shift variable  $z_t$ , the source of uncertainty in the model, evolves (in logs) as an autoregressive process. To operationalize the model, convenient forms of the utility and production functions are assumed and values of the parameters  $\rho$ ,  $\beta$ ,  $\delta$ ,  $\sigma^2$  are obtained by calibration; there is an implicit reference to a particular economy, e.g. the US economy in recent decades.

The question about the welfare implications of a change in the size of technology shocks is interpreted as being about the consequences of a change to the distribution of the technological shift variable  $z_t$ . The question is answered by running the model and calculating the current achieved level of utility and then running it with the hypothesized value of  $\sigma^2$  and comparing the achieved level of utility with that achieved without the change. The details are different but running an experiment involves the same logical steps as solving and simulating a model like the Klein–Goldberger model or conducting one of Patinkin's market-experiments. There are no exact analogues to Patinkin's individual experiments but it is possible to ask how the components of the model economy function, e.g. by how much would the agent prefer this hypothetical consumption stream to that one?

Kydland and Prescott (1996: 70–1) give a dozen examples of 'well-posed' or 'well-defined' questions. Because the piece is a review of achievements, there are no examples of ill-posed questions. Underlying the experiments are unarticulated autonomy judgements (above section 3): thus the  $\sigma^2$  experiment would fail if, say, the preferences over consumption streams actually reflected experience with a particular technology and changed when the technology changed.

The experiment in which the change to  $\sigma^2$  is studied illustrates the case of asking what happens when? This could be a descriptive or an analytical experiment. We might ask, *à propos* the original value of  $\sigma^2$ , 'what is the quantitative nature of fluctuations induced by technology shocks?'. This can be read as an application experiment, although Kydland and Prescott (1996) classify it as a 'developing theory' experiment. Such experiments are considered in the next section.

## 6 'DEVELOPING THEORY' AND STATISTICAL INFERENCE

We have concentrated on the way models in the three approaches are used to answer questions about matters 'outside' the model and have not yet discussed the other type of experiment characterized in section 2 above where the aim of the question is to challenge the model or develop the theory. Much of the inter-approach dispute has been about such experiments, in particular about the test experiment – what it shows and indeed whether it is needed. In the conduct of model-challenging

experiments there is a clear division between the approaches: the VAR and system of equations approaches use statistical inference in model-challenging experiments, the computational experiment does not. It uses a very similar technique for both kinds of experiment while the approaches using statistical inference seem not to. We begin with a sketch of the different approaches to experiments of this second kind and then consider some of the arguments deployed.

The main effort in the Cowles project was directed, not to applying the models, but to constructing them by combining theory with data and to developing the statistical techniques by which this combination could be effected. The project of using data on the outcomes of market experiments to draw conclusions about individual experiments required a new body of statistical inference theory and an investigation of identification, i.e. whether individual experiments can be reconstructed from the results of market experiments. In this approach – and the VAR approach followed – application experiments, like policy simulation, and test experiments, like estimation and testing, seem to be very different. This was the case in the Klein–Goldberger era but it is a matter of appearance, the product of technical limitations, for it is possible to set up the test experiments so they resemble application experiments. This is transparent in a Bayesian treatment where the estimation and testing are done in the same framework as the policy simulation and forecasting. Recent work on estimating real business cycle models, e.g. Smith (1993), has involved running a computational experiment as part of the estimation process has closed another gap.

In the computational experiment approach the estimates (or calibrations) are obtained from other sources, usually from studies using micro-data; the use of calibration is often treated as the distinguishing mark of the approach and it has naturally attracted the attention of methodologists – see e.g. Hoover (1995) and Boumans (2002). The difference between the computational experiment and the Cowles and VAR schemes is not that one uses experiments and the other does not but in the information used to construct the model economy. In the Cowles and VAR approaches the model is estimated and evaluated using macroeconomic data of the kind to which the model is claimed to be relevant while the computational experiment relies on extraneous estimates. Criticism of the computational experiment has concentrated on the question of whether the theory is strong enough to apply in the present instance, where there is adequate empirical support for the model given that it is not evaluated by conventional means. Some of the articles in Hartley *et al.* (1998) argue this case, while others build bridges between the computational experiment and more conventional econometrics.

Kydland and Prescott (1996: 70) introduce the second type of study – ‘developing theory’ – as follows:

These questions typically ask about the quantitative implications of theory for some phenomena. If the answer to these questions is that the predictions of theory, match the observations, theory has passed that particular test. If the answer is that there is a discrepancy, a deviation from theory has been documented.

This is a relaxed post-falsificationist world in which tests are passed but not failed, instead deviations are documented. In Kydland and Prescott's earlier work rejection was a possibility. When they constructed a model (Kydland and Prescott 1982) to 'explain the cyclical variances of a set of economic time series' for the US economy they found the fit 'surprisingly good in light of the model's simplicity and the small number of free parameters' (Kydland and Prescott 1982: 1345). This test experiment was not designed in accordance with the usual statistical principles, however,

We chose not to test our model against the less restrictive vector autoregressive model. This most likely would have resulted in the model being rejected, given the measurement problems and the abstract nature of the model. Our approach is to focus on certain statistics for which the noise introduced by approximations and measurement errors is likely to be small relative to the statistic. Failure of the theory to mimic the behavior of the post-war US economy with respect to these stable statistics with high signal-noise ratios would be grounds for its rejection.

Although the conventional test of over-identifying restrictions is not used, this is a test experiment in the sense of section 2. In preferring a test based on mimicking behaviour Kydland and Prescott (1982) make two points against conventional econometric methods: 'measurement problems' and the 'abstract nature of the model'. The first at least translates into conventional statistical terms while the second may be connected to measures like Akaike's information criterion which consider the number of free parameters in a model; certainly the character of the model would appear on the informal balance sheet of model evaluation.

Kydland and Prescott have since given other grounds for rejecting the use of statistical inference. Identifying the domain of statistical inference with the material experiment, they argue

nature seldom, if ever, provides those who study business cycles with well-designed experiments, and it is not feasible to run controlled experiments with national economies. That is why the tools of R. A. Fisher and of Neyman and Pearson, which are useful in drawing empirical (inductive) inferences, are of little use in business cycle research and why theoretical (deductive) inference now dominates business cycle research.

(Kydland and Prescott 1997: 211)

Kydland and Prescott (1996: 72, 83) also reject testing as incompatible with their notion of theory. Following Lucas (1980) they take ‘theory’ as ‘an explicit set of instructions for building an imitation economy to address certain questions and not a collection of assertions about the behavior of the actual economy’. Consequently, ‘statistical hypothesis testing, which is designed to test assertions about actual systems, is not an appropriate tool for testing economic theory’. Kydland and Prescott are not consistent in their use of this interpretation of ‘theory’ for they also write about assertions which are ‘implications’ of a theory. The system of equations and VAR modellers have generally taken the ‘assertions’ view of theory but if they switched to the ‘instructions’ they describe the tests they perform as tests of the usefulness of the instructions. The choice between these views does not settle the issue of whether one does statistical testing.

The most interesting of Kydland and Prescott’s objections to model fitting are based on their code of question and answer (see section 2 above). They apply this as follows (1996: 73):

The selection and construction of a particular model economy should not depend on the answer provided. In fact, searching within some parametric class of models for the one that best fits a set of aggregate time series makes little sense. For example, if the question is of the type, ‘how much of fact  $X$  is accounted for by  $Y$ ’, then choosing the parameter values in such a way as to make the amount accounted for as large as possible according to some metric is an attempt to get a particular – not a good – answer to the question.

The first sentence takes a principle appropriate to the case where the credibility of the answer is based on the credibility of the model and applies it to the case of fitting a model where we have the answer (the ‘set of aggregate time series’) and are looking for a model that will yield that answer. However the two situations are not governed by the same principles. Their example of ‘how much of fact  $X$  is accounted for by  $Y$ ’ is directed at the analysis of variance of regression analysis and the innovation accounting of VAR analysis. Behind these procedures *are* ways of justifying them as providing ‘good’ answers; depending on the statistical persuasion of the user, the justifications are found in repeated sampling properties or on their following from Bayesian principles.

We have picked out only a few issues raised by experiments that challenge the model, although arguments about such experiments have been central to the debate between the approaches; e.g. Sims began his critique of the systems of equations approach by arguing that the test experiments done under its auspices did not establish what they were thought to establish for the models were not identified. An adequate treatment of the issues of identification and inference would greatly prolong the present discussion,

especially were we to take into account the many hybrids produced in the twenty-five years in which the approaches have coexisted.

## 7 SUMMARY

Starting from Kydland and Prescott's broad notion of running an experiment we have examined the role of the experiment and the rules for its conduct in three styles of macroeconomic modelling, based on the system of equations, the VAR and the computational experiment. These approaches, one from the 1940s and two from the 1980s, continue to inspire work in macroeconometrics. Although Kydland and Prescott have virtually patented the term *experiment*, experimental reasoning in their sense of experiment is essential to all three forms of macroeconometrics. Together the three provide a rich variety of experimental reasoning. Two of the approaches use economic theory but the VAR does not; two of the approaches uses statistical inference but the computational experiment does not. The literature on the computational experiment has generated the most sweeping principles, that on the system of equations the most detailed analyses and that on the VAR the most disputed choices.

We have found some general principles of experimentation in the literature on the three approaches. These include: 'If the theory is strong and the measurements good, we have confidence that the answer for the model economy will be essentially the same as for the actual economy' and 'The selection and construction of a particular model economy should not depend on the answer provided' (Kydland and Prescott 1996: 73 and 83); experiments should be performed that are 'economically meaningful or interesting' (Koop *et al.* 1996: 120). These principles sound like platitudes but applying them requires care and can easily generate disagreement. In our investigation of the practice associated with the different approaches we used a number of distinctions, between model-applying experiments and model-challenging experiments, between descriptive experiments and analytical experiments and between the different senses that have been given to the meaningful and interesting.

The bulk of the discussion above (sections 3–5) has concerned application experiments and has concentrated on the code governing the design of experiments within an approach; there has been a particularly lively discussion of this issue in the VAR school. Most of the between-approach debate has been about the credibility of the models used for applications experiments, and not about the experiments themselves. In section 6 we looked at the use of test experiments to establish the credibility of models and the controversy surrounding it. In this instance the experimental perspective produces a surprising view. Although the experiment is more emphasized in the writing of Kydland and Prescott than in the other literatures of macroeconometrics, the experiment actually has a smaller role

in their project than in others where running an experiment is part of the process by which a particular model economy is selected and constructed.

*John Aldrich*  
University of Southampton  
johnaldrich@soton.ac.uk

*Anna Staszewska*  
University of Łódź  
emfans@uni.lodz.pl

## ACKNOWLEDGEMENTS

We are grateful to the referees for comments and suggestions.

## REFERENCES

- Adelman, I. and Adelman, F. L. (1959) 'The dynamic properties of the Klein–Goldberger model', *Econometrica* 27: 596–625.
- Aldrich, J. (1989) 'Autonomy', *Oxford Economic Papers* 41: 15–34.
- Aldrich, J. (1993) 'Cowles exogeneity and CORE exogeneity', University of Southampton Department of Economics Discussion Paper 9308.
- Aldrich, J. (1995) 'Correlations genuine and spurious in Pearson and Yule', *Statistical Science* 10: 364–76.
- Aldrich, J. (1997) 'Multiple regression grumbles', in D. Conniffe (ed.) *Roy Geary: Irish Statistician*, Dublin: Oak Tree Press, pp. 153–77.
- Aldrich, J. (2005) 'Fisher and regression', *Statistical Science* 20: 401–17.
- Bernanke, B. S. (1986) 'Alternative explanations of the money income correlation', *Carnegie–Rochester Conference Series on Public Policy* 25: 49–100.
- Blanchard, O. J. and Quah, D. (1989) 'The dynamic effects of aggregate demand and supply disturbances', *American Economic Review* 79: 655–73.
- Boumans, M. (2002) 'Calibration of models in experiments', in L. Magnani and N. Nersessian (eds) *Model-Based Reasoning: Science. Technology Values*, New York: Kluwer, pp. 75–94.
- Boumans, M. (2003) 'How to design Galilean fall experiments in economics', *Philosophy of Science* 70: 308–29.
- Boumans, M. and Morgan, M. S. (2001) 'Ceteris paribus conditions: materiality and the application of economic theories', *Journal of Economic Methodology* 8: 11–26.
- Cartwright, N. (1989) *Nature's Capacities and their Measurement*, Oxford: Oxford University Press.
- Christiano, L. J., Eichenbaum, M. and Evans, C. L. (1999) 'Monetary policy shocks: what have we learned and to what end?', in J. B. Taylor and M. Woodford (eds) *Handbook of Macroeconomics*, Amsterdam: North-Holland, pp. 65–148.
- Cooley, T. F. and LeRoy, S. (1985) 'Atheoretical macroeconomics: a critique', *Journal of Monetary Economics* 16: 283–308.
- Ericsson, N. R. and Irons, J. S. (eds) (1994) *Testing Exogeneity*, Oxford: Oxford University Press.

- Fisher, R. A. (1925) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.
- Fisher, R. A. (1935) *The Design of Experiments*, Edinburgh: Oliver & Boyd.
- Frisch, R. (1938) *Autonomy of Economic Relations*, League of Nations memorandum, reproduced in D. F. Hendry and M. S. Morgan (eds) (1995) *Foundations of Econometric Analysis*, Cambridge: Cambridge University Press.
- Geary, R. C. (1963) 'Some remarks about relations between stochastic variables: a discussion document', *Review of the International Statistical Institute* 31: 163–81.
- Goldberger, A. S. (1959) *Impact Multipliers and Dynamic Properties of the Klein–Goldberger Model*, Amsterdam: North-Holland.
- Haavelmo, T. (1944) 'The probability approach in econometrics', *Econometrica* 12(Supplement): i–viii, 1–118.
- Hartley, J. E., Hoover, K. D. and Salyer, K. D. (eds) (1998) *Real Business Cycles: A Reader*, London: Routledge.
- Hausman, D. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.
- Hoover, K. D. (1995) 'Facts and artifacts: calibration and the empirical assessment of real-business-cycle models', *Oxford Economic Papers* 47: 24–44.
- Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan.
- Klein, L. R. (1950) *Economic Fluctuations in the US 1921–41*, New York: Wiley.
- Klein, L. R. and Goldberger, A. S. (1955) *An Econometric Model of the United States 1929–1952*, Amsterdam: North-Holland.
- Koop, G., Pesaran, M. H. and Potter, S. M. (1996) 'Impulse response analysis in nonlinear multivariate models', *Journal of Econometrics* 74: 119–47.
- Koopmans, T. C. (ed.) (1950) *Statistical Inference in Dynamic Econometric Models*, New York: Wiley.
- Koopmans, T. C., Rubin, H. and Leipnik, R. B. (1950) 'Measuring the equation systems of dynamic economics', in T. C. Koopmans (ed.) *Statistical Inference in Dynamic Econometric Models*, New York: Wiley, Ch. II.
- Kydland, F. E. and Prescott, E. C. (1982) 'Time to build and aggregate fluctuations', *Econometrica* 50: 1345–70.
- Kydland, F. E. and Prescott, E. C. (1991) 'The econometrics of the general equilibrium approach to business cycles', *Scandinavian Journal of Economics* 93: 161–78.
- Kydland, F. E. and Prescott, E. C. (1996) 'The computational experiment: an econometric tool', *Journal of Economic Perspectives* 10: 69–85.
- Kydland, F. E. and Prescott, E. C. (1997) 'Correspondence: response to Friedman', *Journal of Economic Perspectives* 11: 209–12.
- Leontief, L. (1937) 'Implicit theorizing: a methodological criticism of the neo-Cambridge school', *Quarterly Journal of Economics* 51: 337–51.
- Levtchenkova, S., Pagan, A. R. and Robertson, J. C. (1998) 'Shocking stories', *Journal of Economic Surveys* 12: 507–32.
- Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.
- Lucas Jr, R. E. (1976) 'Econometric policy evaluation: a Critique', in K. Brunner and A. Meltzer (eds) *The Philips Curve and Labor Markets*, Vol. 1 of the *Carnegie–Rochester Conferences on Public Policy*, a supplementary series to the *Journal of Monetary Economics*, Amsterdam: North-Holland, pp. 19–46.
- Lucas Jr, R. E. (1980) 'Methods and problems in business cycle theory', *Journal of Money, Credit and Banking* 12(4): 696–715.
- Lucas Jr, R. E. and Sargent, T. J. (eds) (1981) *Rational Expectations and Econometric Practice*, London: George Allen and Unwin.

- Machamer, P., Darden, L. and Craver, C. F. (2000) 'Thinking about mechanisms', *Philosophy of Science* 67: 1–25.
- Mäki, U. (2005) 'Models are experiments, experiments are models', *Journal of Economic Methodology* 12: 303–15.
- Marschak, J. (1949) 'Statistical inference from non-experimental observations: an economic example', in *Proceedings of the International Statistical Conferences*. International Statistical Institute, Vol. III, pp. 289–98.
- Marschak, J. (1950) 'Introduction', in T. C. Koopmans (ed.) *Statistical Inference in Dynamic Econometric Models*, New York: Wiley, Ch. I.
- Morgan, M. S. (1990) *History of Econometric Ideas*, Cambridge: Cambridge University Press.
- Morgan, M. S. (2002) 'Model experiments and models in experiments', in L. Magnani and N. Nersessian (eds) *Model-Based Reasoning: Science, Technology Values*, New York: Kluwer, pp. 41–58.
- Morgan, M. S. (2005) 'Experiments versus models: new phenomena, inference and surprise', *Journal of Economic Methodology* 12: 317–29.
- Patinkin, D. (1956) *Money, Interest and Prices: An Integration of Monetary and Value Theory*, Evanston, IL: Harper & Row.
- Samuelson, P. A. (1947) *Foundations of Economic Analysis*, Cambridge, MA: Harvard University Press.
- Samuelson, P. A. (1976) 'The optimum growth rate for population: agreement and evaluations', *International Economic Review* 17: 516–25.
- Sargent, T. J. (1976) 'The observational equivalence of natural and unnatural rate theories of macroeconomics', *Journal of Political Economy* 84: 631–40.
- Schumpeter, J. A. (1954) *History of Economic Analysis*, London: Allen & Unwin.
- Sims, C. A. (1980) 'Macroeconomics and reality', *Econometrica* 48: 1–48.
- Sims, C. A. (1981) 'An autoregressive index model for the U.S. 1948–1975', in J. Kmenta and J. B. Ramsey (eds) *Large-Scale Macro-Econometric Models*, Amsterdam: North-Holland, pp. 283–327.
- Sims, C. A. (1982) 'Policy analysis with econometric models', *Brookings Papers on Economic Activity* 1: 107–52.
- Smith, A. A. (1993) 'Estimating nonlinear time-series models using simulated vector autoregressions', *Journal of Applied Econometrics* 8: S63–S84.
- Staszewska, A. (2004) 'Inference in VAR modelling: some preliminaries', in W. Welfe and A. Welfe (eds) *Macromodels 2003*. Proceedings of the 30th International Conference Warsaw, 3–6 December, Wydawnictwo Uniwersytetu Łódzkiego, pp. 199–216.
- Whitaker, J. K. (1991) 'Ceteris paribus', in J. Eatwell, M. Milgate and P. Newman (eds) *The New Palgrave: A Dictionary of Economics*, London: Macmillan, pp. 396–97.
- Woodward, J. (2002) 'What is a mechanism? A counterfactual account', *Philosophy of Science* 69: S366–S377.
- Yule, G. U. (1899) 'An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades (Part I)', *Journal of the Royal Statistical Society* 62: 249–96.