1

Introduction

by Lars Peter HANSEN and Thomas J. SARGENT

We doubt that the material in this book could be described as "classic", but so much of it has been 'underground' for so long that we seized the opportunity to publish it in the "Underground Classics" series.[1] The papers in this volume were written over a period of about twelve years, with versions of most of the papers being written between 1979 and 1982. These papers report some of our efforts to make the hypothesis of rational expectations econometrically operational. We delayed publishing these papers for one reason or another, mostly because we believed that some of the arguments could be improved or because we were too busy with other projects to put the finishing touches on these papers. We welcome the opportunity to publish these papers, rough as some of them are, in the "Underground Classics" series, whose editor Spencer Carr has told us that polish and finish would only detract from the underground flavor.

We are macroeconometricians, and have come at the hypothesis of rational expectations from the perspective of macroeconometrics. In the last two decades, the hypothesis of rational expectations has invaded economics from a variety of sources, including game theory and general equilibrium theory. But in macroeconomics, the first invaders were time series econometricians, who in the late 1960's were seeking methods of restricting the parameters of lag distributions in their econometric models.[2] In the late 1970's, the focus of attention changed from restricting distributed lags to restricting vector autoregressions. The papers in this volume are intended to carry on and contribute to this tradition.

For macroeconometricians, an exciting aspect of rational expectations models is the way their solution (or equilibrium) restricts an entire stochastic process of observables. Subsumed within the restrictions are explicit models of the various sources of 'disturbances' or errors in economic relations, aspects of econometric models that drive estimation and inference, but about which pre-rational expectations (mostly non-stochastic) macroeconomic theories had typically been silent. Rational expectations modelling promised to tighten the link between *theory* and *estimation*, because the objects produced by the theorizing are exactly the objects in terms of which econometrics is cast, e.g., covariance generating functions, Markov processes, and ergodic distributions.

Work on rational expectations econometrics has divided into two complementary but differing lines. The first line aims more or less completely to characterize the restrictions that a model imposes on a vector stochastic process of observables, and to use those restrictions to guide efficient estimation. This line is a direct descendant of the full system approach to estimating simultaneous equations models. It aims to estimate all of the *deep parameters* of a model simultaneously by exploiting the cross-equation restrictions that a rational expectations model imposes on those parameters. The benefits of this line of attack are described by Hansen and Sargent (1980a) and Sargent (1981), principal ones being the following three: (i) the ability to handle a range of assumptions about stochastic error processes and about decision variables that are present in the model but missing in the econometrician's data set (e.g., effort); (ii) the promise of estimating the full range of parameters required to study responses to various policy interventions; and (iii) the econometric advantages of increased efficiency that are associated with full system estimation methods.

The second line of work is the application of method of moments estimators to estimating the parameters that appear in the Euler equations associated with dynamic optimum problems. Hall (1978) and Hansen and Singleton (1982) recognized that if strong assumptions are made about those processes in an Euler equation that are unobservable to the econometrician (most often, that they are not present), then an Euler equation implies that a set of orthogonality conditions hold for the econometrician's data set. These orthogonality conditions can contain enough information to identify the parameters in the return function corresponding to that Euler equation. Thus, this approach holds the promise of estimating some parameters without the need to estimate (or indeed even to specify) a complete equilibrium model.

The second line of work has the advantage that it is easier to implement than the first. Its disadvantages are the restrictive assumptions on unobservables needed to validate the approach, and the fact that even when it is applicable, it does not attempt to estimate the full range of parameters that are typically required to analyze an interesting range

of policy interventions. This second approach has caught on more than the first in applied work, undoubtedly mainly because of its ease.

Most of the papers in this book are contributions to the first line of work (though the third, fourth, and fifth chapters can be viewed as embracing a strategy that is somewhere in between the two lines). All of the papers provide ways of interpreting and restricting *vector autoregressions* in the light of some version of a rational expectations or equilibrium model. The economic theories used in this book are without exception linear rational expectations models. We chose this class purposefully because they match up so naturally with vector autoregressions. Prices and quantities in these models are determined by the interactions of agents who are optimally responding to disturbances informing them about their current and future prospects. Underlying several of the papers in this volume is a recurring theme about the link between the innovations in a vector autoregression and the disturbances to agents' information sets, a theme to which we now turn.

The Multitude of Moving Average Representations

Chapter 2 by Hansen and Sargent describes elements of the theory of linear least squares prediction. This chapter is just a set of lecture notes we have used to teach least squares theory to our students. Much of the discussion in the remainder of the book is cast in terms of the objects defined in this theory. Linear least squares prediction theory studies a vector stochastic process by decomposing it into two orthogonal pieces: a part that can be forecast as a linear function of its own past values, and a part that cannot be forecast. A key construction of the theory is Wold's decomposition theorem. One starts with an $(n \times n)$ positive semi-definite matrix sequence $\{C_x(\tau)\}$ and interprets $C_x(\tau)$ as the covariance $Ex_t x'_{t-\tau}$ for an $n \times 1$ vector stochastic process $\{x_t\}$. From the information in $\{C_x(\tau)\}$, one constructs a sequence of *projections* or *autoregressions* of x_t onto a linear space spanned by $(x_{t-1}, \ldots, x_{t-n})$. By studying the behavior of these projections as $n \to \infty$, one arrives at a decomposition that expresses x_t as the sum of the part that can be predicted linearly from past values and an *innovation* u_t that is orthogonal to the predictable part. By construction, the innovation process u_t is serially uncorrelated and lies in the space spanned by current and lagged x's (i.e., it is a forecast error). Under some additional assumptions, x_t can be expressed as a *moving average* of the innovation u_t, namely,

$$x_t = \sum_{j=0}^{\infty} D_j \, u_{t-j} \; . \tag{1}$$

This is known as the *Wold moving average representation* for x_t, and the $\{u_t\}$ process is said to be a *fundamental* white noise process for x_t. The term *fundamental* denotes that u_t is a white noise in terms of which x_t possesses a one-sided moving-average representation, and that u_t lies in the linear space spanned by current and lagged x_t's.

By virtue of the construction that leads to the Wold representation, the $\{u_t\}$ process is a white noise that corresponds to an innovation in an infinite order vector autoregression. Thus, the "innovation accountings" of Sims (1980) are statements about a moving-average representation (1).

Moving average representations are not unique for two distinct reasons. First, one can always multiply u_t in (1) by a nonsingular matrix U and obtain another Wold moving-average representation, namely,

$$x_t = \sum_{j=0}^{\infty} (D_j \, U^{-1}) \; (U u_{t-j}) \; , \tag{2}$$

in which $(U u_t)$ is the new innovation and $(D_j \, U^{-1})$ is the new impulse response function. This kind of nonuniqueness is the type confronted by Sims (1980) in his discussion of alternative orthogonalization schemes for the innovation process. It leaves the optimal predictions implied by all of the associated Wold moving average representations and the information in the vector $U u_t$ unaffected. Since U is nonsingular, the history of $U u_t$'s spans exactly the same linear space as does the history of u_t's (or the history of x_t's).

The second type of nonuniqueness of (1) does affect the information content of the residuals. There exists a family of other moving average representations

$$x_t = D^* \, (L) u_t^* \; , \tag{3}$$

where

$$D^* \, (L) = \sum_{j=0}^{\infty} D_j^* \, L^j \; , \quad \sum_{j=0}^{\infty} \text{ trace } \; D_j^* \; D_j^{*\prime} < +\infty \; ,$$

and where $D^* \, (L)$ satisfies

$$D(z) \, E u_t \; u_t' \, D(z^{-1})' = D^*(z) \; E u_t^* \; u_t^{*\prime} \, D^* \, (z^{-1})' , \quad |z| = 1 \; . \tag{4}$$

In (3), $\{u_t^*\}$ is an $(m \times 1)$ white noise, where $m \geq n$, and $D^* \, (L)$ is an $n \times m$ matrix polynomial in the lag operator L. Given $D(z)$, any $D^*(z)$ that satisfies (3) defines a moving average representation with associated vector white noise $\{u_t^*\}$. For *most* such representations, namely all non Wold representations, the history of u_t^* spans a *larger* space than the space generated by the corresponding history of x_t's. In general, (3) is a representation in which the history of x_t's fails to *reveal* the corresponding history of u_t^*'s.

Much of this book involves studying settings in which an economic model has an equilibrium that is most naturally represented in the form of a moving average representation of the type (3), but in which the internal structure of the model implies that this representation is not automatically a Wold representation. This poses a problem for interpreting the innovations u_t in a vector autoregression in terms of the innovations u_t^* that occur in the economic model. Some of the papers in this volume are concerned with characterizing the dimensions of this interpretation problem within particular concrete contexts. Other papers are concerned with providing methods for circumventing the problem conceptually and econometrically.

Exact Linear Rational Expectations Models

The paper "Exact Linear Rational Expectations Models" studies a class of models in which we immediately have to confront the multiplicity of moving average representations associated with a given stationary stochastic process. This paper studies a special class of linear rational expectations models that are constructed out of two sorts of relationships. Economic theory enters only in the first set of relationships, which consist of a set of linear relationships involving *only* current and past values of a subset of variables x_t observable by the econometrician, and expectations of future values of those observable variables. The qualifier "only" defines what we mean by an *exact* model – there are no disturbance terms in this relationship from econometrician's point of view. The fact that the econometrician has access to a smaller information set than do the agents is the only possible source of an econometric error term in this relationship. The second set of relationships are informational in nature, being the piece of a complete moving-average representation that governs the subset of observables x_t being forecast by the agents in the model. The strategy in this paper is to deduce the restrictions on the moving average of the *entire* vector x_t implied by the hypothesis of rational expectations. It turns out to be easy to write down these restrictions, and to describe strategies for estimating moving average representations subject to them.

This model formulation strategy naturally poses the following question: given that the model is correct, to which of multiplicity of moving-average representations do the rational expectations restrictions apply? In answering this question, we in effect characterize the extent of the problem of identifying moving-average representations consistent with an *exact* rational expectations model. This characterization has implications for likelihood-based procedures for estimation and inference because it indicates that the likelihood function will have multiple peaks corresponding to different moving-average representations. Despite this identification problem, the restrictions implied by these models are testable because the underidentified aspects pertain only to the flow of new information and not to the way the information restricts the covariances across variables.

Section 4 of "Exact Linear..." has some curiosity value. It introduces nonstationarities in the time series and then demonstrates the sense in which the variables are *cointegrated*, although we did not originally apply that term because we were unaware of the work of Granger and co-workers at the time that our paper was initially drafted.

The solution strategy employed in "Exact Linear..." has much in common with that used by Whiteman (1983). Whiteman extensively explores how to impose the rational expectations restrictions on the moving average representations for models that are *fully specified*, in the sense that as many relationships are specified as variables that are determined by the model. In contradistinction our exact linear rational expectations models are incomplete in that the laws of motion for the information variables and variables being forecast are not completely specified. Despite this difference in economic interpretation between our work and Whiteman's, Whiteman's work shows that many of the mathematical methods used in "Exact Linear..." can be useful in formulating complete models.

The class of models in "Exact Linear..." has a number of applications, but is quite special in that the theory must come in the form of an *exact* model, which means that the econometric model can contain no source of error *except* that the econometrician conditions on a smaller information set in forming forecasts than do economic agents.[3] The next chapter, "Two Difficulties in Interpreting Vector Autoregressions", discusses a more general setting in which there are additional sources of disturbances in econometric relations.

Two Difficulties in Interpreting Vector Autoregressions

One common case in which the version of (3) delivered by economic theory fails to match up with the Wold representation occurs when the number of shocks m impinging on agents' information exceeds the number n of processes observed by the econometrician. In this case, the u_t's are bound to summarize and confound the effects of the u_t's. However, the problem can emerge even when $m = n$. The paper "Two Difficulties in Interpreting Vector Autoregressions" studies the problem in two distinct contexts in which $n = m$. The first context is that of a dynamic equilibrium model in which two shocks are impinging on agents' information sets, namely, *supply* and *demand* or *endowment* and preference shocks. The equilibrium of the model is represented as a stochastic process for price and quantity that is a moving average of the shocks in agents' information sets (i.e., the u_t's). The paper studies how these shocks are related to the innovations in a vector autoregression for price and quantity. The paper describes how to express the u_t's as distributed lags of the u_t^*'s. When these distributed lags are not concentrated at zero lag, the $\{u_t\}$ process contains less information than does the $\{u_t^*\}$ process.[4]

The second context studied in the "Two Difficulties..." chapter is that of aggregation over time. The problem here is that the econometrician has data sampled at a coarser interval than that of economic agents. Taking this idea to the limit, suppose that economic agents receive and process information in continuous time, and that an economic theory is in the form of a continuous time version of (3). (Think of a limiting version of (3) approached by successively reformulating (3) at finer and finer sampling intervals). Under what circumstances will the impulse response functions from the vector autoregression associated with a discrete time, sampled version of the x_t process resemble in shape the impulse response function in continuous time? This is a version of the aggregation over time question of Sims (1971) and Geweke (1978), who studied how well discrete time distributed lags would approximate their underlying continuous time counterparts. Rather than studying how these distributed lags (projections of one variable on leads and lags of another) match up, we study how the moving-average representations match up. The paper describes some conditions on the continuous time stochastic process involving *smoothness* (i.e., mean square continuity and differentiability) under which a discrete time moving average must fail to match up well with the underlying continuous time one.

The "Two Difficulties..." chapter assumes that the underlying continuous time model takes the form of a rational spectral density, which automatically imposes continuity on the kernel defining the continuous time moving average representation. In Chapter 10, which was written by Albert Marcet, this continuity requirement is dropped. Marcet studies discontinuous continuous time moving average kernels, in particular, how they affect the closeness of the continuous and discrete time moving average representations. Marcet offers a useful characterization that allows him to create a number of interesting examples in which the continuous and discrete time moving-average representations diverge.

A failure of the moving averages associated with vector autoregressions to match up with those associated with an underlying economic theory suggests caution in interpreting innovation accountings based on estimates of a theoretical vector autoregressions. If a researcher is willing to impose sufficient economic theory on the process of econometric estimation, the difficulties described above can be overcome in the sense both that consistent and efficient parameter estimates can be obtained, and that estimates of the moving-average representation corresponding to the economic model can be computed. In discrete time, Hansen and Sargent (1980a, 1981a) described strategies for carrying out such estimation. Corresponding methods for continuous time were described by Hansen and Sargent in unpublished papers (1980b, 1981d). Versions of these continuous time methods are described in this volume in papers by Hansen, Heaton, and Sargent and by Hansen and Sargent.

Three Papers on Continuous Time Rational Expectations Models

Chapters 7, 8, and 9 study three problems that must be solved if a continuous time dynamic equilibrium model is to be estimated via maximum likelihood methods. Chapter 7 by Hansen, Heaton, and Sargent and chapter 8 by Hansen and Sargent are concerned with computing the solution of a continuous time model in a form designed for econometric tractability. These two papers exploit the *certainty equivalence* feature of linear dynamic models, namely, that their solution can typically be broken into two separate steps: "optimization" and "forecasting". In effect, Hansen, Heaton, and Sargent describe how the equilibrium of a class of deterministic continuous time linear equilibrium models can be computed by solving a linear quadratic optimal control problem. The paper describes how a fast algorithm (a matrix sign algorithm) can be put to work on this problem. The product of Hansen, Heaton, and Sargent's calculations is an equilibrium in *feedback part, feedforward part* form. This is the solution of a deterministic version of the model, in which future paths of forcing variables are known with certainty. To obtain the solution for stochastic versions of the model in which the forcing functions are stochastic processes, one simply substitutes linear least squares forecasts for future values in the feedforward part, leaving the feedback part unaltered. In chapter 8, Hansen and Sargent describe a set of convenient formulas for computing the feedforward parts in the stochastic case. These are the continuous time counterparts of formulas described by Hansen and Sargent (1980a, 1981b) for the discrete time case.

By combining the results in these two papers, one obtains a representation of the solution in the form of a vector linear stochastic differential equation. This representation provides a continuous time version of the solutions described by Hansen and Sargent (1980a, 1990). Given such a representation, standard methods can be used to compute the likelihood function of the continuous time model conditioned on discrete time data. These methods are described by Jones (1980), Ansley and Kohn (1983), Bergstrom (1983), Harvey and Stock (1985), and Hansen and Sargent (1980b). Christiano, Eichenbaum, and Marshall (1990) have applied such methods to study consumption smoothing models using U.S. time series data.

Chapter 9 by Hansen and Sargent treats a special case of the identification problem that must be solved in estimating a continuous time model from discrete time data. As described by Phillips (1973) and Hansen and Sargent (1981c), without some restrictions on the continuous time model, there is typically a multitude of continuous time models that is consistent with a given discrete time covariance generating function. This is a version of the classic *aliasing* phenomenon. Hansen and Sargent illustrate how the cross-equation restrictions imposed by rational expectations can serve to resolve this identification problem. While Hansen and Sargent obtain analytical results only for a special case, their results suggest numerical methods that can be used to check identification in more general models.

Testing Present Value Budget Balance

Chapter 5 by Hansen, Roberds, and Sargent describes a class of models in which the moving-average representation delivered by theory is not a Wold representation. Research on the subject of this paper was initiated in response to a question posed by Robert E. Lucas, Jr. at a conference in October 1985. Lucas asked what restrictions would imposed on a joint stochastic process describing net of interest government expenditures and taxes by the assumption of present value budget balance. Lucas conjectured that even with a constant real interest rate, the restriction would be a weak one because of the ability to postpone repayment now via an indefinite promise to run surpluses later. An incomplete analysis of the issue was made by Sargent (1987b), who showed that the hypothesis of present value budget balance imposes the restriction that the present value of the impulse response coefficients of the deficit to each innovation in agents' information set is zero. This is also one of the restrictions imposed by the consumption smoothing model used by Hall (1978) and the tax smoothing model used by Barro (1979). One of Hansen, Roberds, and Sargent's aims is to characterize exhaustively the restrictions implied by a class of models including Hall's and Barro's as special cases.

Let g_t denote government expenditures and let τ_t denote tax collections, both net of interest. Hansen, Roberds, and Sargent start by studying the case in which observations on g_t, τ_t, but *not* on the debt, are available to the econometrician. They assume a constant real interest rate. They begin with the observation that, in general, the restriction that the present value of the moving-average coefficients for the deficit $\tau_t - g_t$ be zero implies that that moving-average cannot be a Wold representation. The reason is that the restriction itself implies that the moving-average polynomial in the lag operator is not invertible. This has the implication that the restriction ought not to be tested by checking whether it holds for the impulse response function associated with a vector autoregression (i.e., a Wold representation).

Hansen, Roberds, and Sargent go on to show that the restriction itself is vacuous unless additional restrictions are imposed on the moving-average representation. In particular, they show that given a moving-average for $\{(g_t, \tau_t)\}$ that violates the restriction, one can always find another moving-average representation that satisfies the restriction. To build this alternative representation, one just needs sufficient flexibility in the parameterization of the alternative moving-average representation. This result confirms Lucas's initial skepticism about the restrictiveness of the budget balance restriction.

Hansen, Roberds, and Sargent then describe two contexts in which the restriction is testable. The first context, described in sections 3-5 of their paper, imposes additional theoretical structure in the form of a version of an optimal consumption smoothing model. The second assumes that additional data are available, namely, a time series on the stock of real interest bearing debt.

In the first setting, Hansen, Roberds and Sargent study whether the hypothesis of present value budget balance adds any testable implications to a martingale model for the marginal utility of consumption. They show that it does add a testable restriction, and that a form of this additional restriction continues to hold in the case in which preferences are nonseparable in consumption. Hansen, Roberds, and Sargent show that the source of this testable restriction is the ability that the martingale model has to identify one component of agents' information set, namely, the innovation to *consumption outlays*, which the theory states is a linear combination of the innovations to agents' information sets.

In sections 4 and 5 of their paper, Hansen, Roberds and Sargent describe and implement a strategy for testing this implication of present value budget balance in the context of a martingale model with nonseparable preferences. Hansen, Roberds, and Sargent describe a *semi-parametric* testing strategy, in that they do not directly specify and estimate the various technology and preference parameters in their underlying model (as would be done, for example, if one were pursuing the estimation strategy for those models described in Hansen and Sargent 1990). Instead, they parameterize some particular lag distributions that are mongrel parameterizations of the structural parameters. This strategy is motivated by a desire to focus on the present value budget restriction while remaining noncommittal about details of the nonseparabilities and the number of goods in the model. Hansen, Roberds, and Sargent apply this test to U.S. data on consumption and labor income, and find little evidence against the present value budget restriction. It might be interesting to repeat this test for a generalized *tax smoothing* model using U.S. data.

However, Section 6 raises a cautionary note concerning the interpretation of these tests. Section 6 establishes the existence of a sequence of false models that satisfy the restrictions but that approximate data that violate the restriction arbitrarily well. This result has to dampen somewhat one's enthusiasm for our semi-parametric strategy, since it suggests that if we get nonparametric enough, the present value budget restriction becomes virtually vacuous even with the martingale model also imposed.[5]

Section 7 of the Hansen, Roberds, Sargent paper changes the setup in two ways, first to permit time varying interest rates and second to add observations on debt to the expenditure and revenue series. The paper shows that the present value budget balance restriction leads to an exact linear rational expectations model. In a separate paper by Roberds, such a model is estimated and the restriction is tested for U.S. data on government expenditures and taxes. Roberds finds evidence against the restriction for these data.

Notes

1. Chapter 2 was written in 1981–82. Chapter 3 was written in 1980–81 (Hansen and Sargent 1981e), and revised in 1990. Chapter 4 was written in 1982, with minor revisions being made in 1984 (Hansen and Sargent 1984) and 1989. Most of chapter 5 was written in 1987, with the empirical work being completed in 1990. Chapter 6 was written in 1988. Chapters 7 and 8 were written in 1988–90. They amount to revisions and extensions of ideas that initially appeared in working papers by Hansen and Sargent that appeared in 1980 and 1981 (Hansen and Sargent (1980b, 1981d). Chapter 9 was completed in 1980–81 (Hansen and Sargent 1981c). Chapter 10 was written as part of Albert Marcet's Ph.D. dissertation, which was completed in 1987.

2. See Muth (1960, 1961), Nerlove (1967), Griliches (1967), Sims (1974), Lucas (1972), and Sargent (1971).

3. This way of introducing errors in econometric models was used to great advantage by Shiller (1972).

4. The paper also describes how its partial equilibrium model is to be interpreted as a special case of the class of general equilibrium models studied by Hansen and Sargent (1990).

5. See Sargent (1987b, chapter XIII) for a discussion of the relationship between Hall's (1978) model and Barro's (1979) tax smoothing model. Evidently, there are tax smoothing models that are similarly related to the general class of consumption smoothing models described by Hansen, Roberds, and Sargent or by Hansen and Sargent (1990).

2

Lecture Notes on Least Squares Prediction Theory

by Lars Peter HANSEN and Thomas J. SARGENT

1. Introduction

In these notes we establish some basic results for least squares prediction theory. These results are useful in a variety of contexts. For instance, they are valuable for solving linear rational expectations models, representing covariance stationary time series processes, and obtaining martingale difference decompositions of strictly stationary processes.

The basic mathematical construct used in these notes is an inner product defined between two random variables. This inner product is calculated by taking the expectation of the product of the two random variables. Many of the results obtained using this particular inner product are analogous to results obtained using the standard inner product on multi-dimensional Euclidean spaces. Hence intuition obtained for Euclidean spaces can be quite valuable in this context as well.

The formal mathematical machinery that is exploited in these notes is the Hilbert space theory. There is a variety of references on Hilbert spaces that should provide good complementary reading, e.g. Halmos (1957) and Luenberger (1969).

2. Prediction Problem

In this section we specify formally the problem of forecasting a random variable y given a collection of random variables H. This problem is sufficiently general to include conditional expectations and best linear predictors as special cases. We also consider a second problem that is closely related to the prediction problem. This second problem is termed the orthogonality problem and can be interpreted as providing a set of necessary and sufficient first-order conditions for the prediction problem. In particular, we will show that these two problems have the