9. Post-orthodox econometrics

Paul Ormerod

I INTRODUCTION

Presenting a paper to any conference with the name 'Keynes' in the title, particularly to one sponsored by the Post Keynesian Study Group, is fraught with danger. For the exegitical debate about 'what Keynes really meant' is often carried out with as much fervour as the arguments between the Fathers of the early Church. Indeed, some of the latter material has a familiar sound. Origen, for example, proposed a theory of Redemption in which the soul fluctuated between Heaven and Hell, a suggestion sternly denounced as heresy by Augustine, who specifically denied the existence of endogenous cycles.

It is as well to make my own position clear from the outset. Keynes is obviously a very important figure in economics. But it is now sixty years since the *General Theory* was written, and the world has moved on. Particularly outside the discipline of economics, scientific methodology has changed. We should be concerned to use advances in knowledge to increase our understanding of the dynamics of capitalism, regardless of whether Keynes – or Smith or Ricardo or Marx, for that matter – really meant the same thing. So I do not intend to enter into definitional arguments about Keynesianism. The aim instead is to review how we should approach applied macroeconomic analysis.

Section II of the chapter considers briefly the post-war research programme in conventional macroeconomic modelling. Both in terms of forecasting accuracy and of structural understanding of the economy, very little progress has been made. Section III points out that the failures of forecasting are inherent in the nature of macroeconomic time series. It is simply not possible with most macro data series to generate forecasts over time which have an error variance significantly less than the variance of the data themselves. An important implication of this result is that conventional short-run stabilization policies cannot work in any meaningful sense, even if the structure of the economy were well understood. In Section IV, I argue that most existing applied macroeconomic relationships are overfitted. In local time neighbourhoods very simple linear models offer a good explanation of key macroeconomic relationships in the OECD economics. However, a small number of shocks have caused permanent shifts in key relationships, such as those between inflation and unemployment and between unemployment and growth. It is the task of political economy and not econometrics alone to understand the reasons for such shifts.

II THE CONVENTIONAL RESEARCH PROGRAMME IN MACROECONOMICS

Most existing macromodels are based upon the synthesis which became 'Keynesian' economics. Depending upon the various strengths of key linkages, these models are capable of generating a wide range of different policy simulation results, from ones close to the position of pure monetarism to ones in which fiscal policy can have a permanent effect on the economy. The criticisms which follow apply equally well, however, to the minority of models which describe themselves in alternative terms, such as 'monetarist' or 'supply-side'.

Such conventional macroeconometric models, whatever their theoretical and empirical leanings, are of little or no value. There are two aspects to discuss here: first, their forecasting record, and, second, their structural understanding of the economy. It is important to note that a great deal of public research money and effort has gone into these models in the postwar period, especially in the past twenty years. Despite this, no real progress has been made using this methodology.

The forecasting record is poor and shows no signs of improving. The OECD (1993), for example, note that over the 1987–92 period one-year ahead forecasts of GDP growth and inflation could not beat the naive rule that next year's growth/inflation will be the same as this year's. These forecasts were carried out by the governments of the G7 countries, the IMF and the OECD itself. As examples of the one-year ahead forecasting record for GDP growth, for the US economy recessions have generally been forecast prior to their occurrence, and the recessions following the 1974 and 1981 peaks in the level of output were not recognised even as they took place (Stekler and Fildes, 1999). Further, growth has generally been overestimated during slowdowns and recessions whilst underestimates occurred during recoveries and booms (Zarnowitz and Braun, 1992). For the UK, the predictions of the Treasury over the 1971–96 period have been at least as good as those of other forecasters, but the mean absolute annual forecast error for these one-year ahead predictions was 1.45 per cent of GDP, compared to an actual mean absolute change of 2.10 per cent (Mellis and Whittaker, 1996). In 13 European countries over the 1971–1995 period, the average absolute error was 1.43 per cent of GDP, compared to the average annual change of 2.9 per cent (Oller and Barot, 1999).

Of course, the raw output of the models is adjusted by the forecasters, but such literature as exists on this topic suggests that the unadjusted forecasts of the models tend to be even worse (see, for example, Clements, 1995, for a list of such references).

Macro models in the UK, thanks to their reliance upon public funding, are the most sophisticated in the world, in the sense of being closest to academic work on economic theory and on econometrics. Yet their policy properties are still far from converging, despite the research effort devoted to them. Church *et al.* (1993) give simple properties of the six leading UK models of 1992 vintage, looking at straightforward questions such as the public spending multiplier. In some ways, the results are even more disparate than those reported for the models of 1977 vintage by Laury *et al.* (1978). On the issue, for example, of the impact on the price level of a change in the VAT rate, the models even give different signs. In some, prices rise, in others, they fall.

A technically more sophisticated exercise which demonstrates the point on policy properties was published recently by Bray *et al.* (1995). The research team used the London Business School, National Institute, Oxford Economic Forecasting and Treasury models of the UK economy to carry out a policy optimization exercise through to the year 2002. The broad objectives were to achieve low inflation, to reduce unemployment, to maximize growth, to keep government borrowing within certain constraints, to stabilize the exchange rate and to meet targets on the balance of payments. The available instruments spanned the range of conventional fiscal and monetary tools, being the standard rate of income tax, government expenditure and the short-term interest rate. The differences in the model results are, as with the simple exercise on a VAT change, not merely quantitative but are different in direction. Compared to the base forecast, for example, the Treasury, LBS and NI models require interest rates to be lower almost throughout the whole period, albeit by widely different amounts, whereas the OEF model requires them to be higher. As a percentage of GDP, government expenditure is higher in the NI model compared to the base forecast, much lower in the LBS model, virtually unchanged in the OEF, and both higher and lower over time in the Treasury model.

In summary, despite an intensive research effort over several decades, conventional macro models are capable neither of producing forecasts whose error variances are significantly less than the variances of the data series being forecast nor of making progress in understanding the impact on the economy of the orthodox instruments of fiscal and monetary policy.

III FORECASTING IN ECONOMICS

Despite the accumulation of evidence, conventional macroeconomic modellers continue to work under the illusion that the application of yet more sophisticated orthodox economic theory and yet more sophisticated econometrics of this kind will enable them both to make better forecasts and to understand the structure of the economy better (see Hall, 1995, for a recent example). In this section I address the question of the inherent forecastability of macroeconomic time-series data and suggest that, in the current state of scientific knowledge, it does not appear possible to generate macro forecasts in which the variance of the forecast error is significantly less than the variance of the data themselves.

The idea that the business cycle is intrinsically unpredictable is not new. Fisher (1925) suggested more than seventy years ago that the business cycle was inherently unpredictable because, in modern terminology, the dimension of the problem is too high relative to the available number of observations. He argued that movements over time in the volume of output were 'a composite of numerous elementary fluctuations, both cyclical and noncyclical', and quoted approvingly from his contemporary Moore, who wrote that '[business] cycles differ widely in duration, in intensity, in the sequence of their phases and in the relative prominence of their various phenomena'. In such circumstances, even though deterministic structure exists, given the limited amount of data available, it would be virtually impossible to distinguish data generated by such a system from data which were genuinely random in terms of their predictability.

Interestingly, David Hendry, the major econometric theoretician of the conventional macroeconomic modelling research programme over the past twenty years, appears to be moving to a view which is close to this. Clements and Hendry (1995) suggest that there is 'a limit on our ability to forecast even with parameter constancy'. They give as their major reason the inherent uncertainty of the model structure. This is a point which Chatfield (1995), a mathematical statistician rather than econometrician, has emphasized for some time. The estimation of model parameters traditionally assumes that the model has a prespecified known form, and takes no account of possible uncertainty regarding the model structure. In particular, this approach postulates implicitly the existence of a single, 'true' model, an assumption which both Clements and Hendry and Chatfield question, the latter writing: 'in practice model uncertainty is a fact of life and is likely to be more serious than other sources of uncertainty which have received far more attention from statisticians'.

Clements and Hendry speculate that 'it may be that only a few periods ahead are necessary for economic stabilisation policy (e.g. 4–8 quarters) and that forecasts are informative over this horizon'. Unfortunately, the evidence suggests that this is probably not the case. In the last few years, for example, statisticians have begun to apply non-linear estimation techniques to macroeconomic data series. Potter (1995) and Tiao and Tsay (1994) investigated quarterly changes in real US GNP from 1947 to 1990, using the first difference of the natural log of GNP. Tiao and Tsay used a threshold autoregressive model in which the data series was partitioned into four regimes determined by the patterns of growth in the previous two quarters. Potter also takes the threshold autoregressive approach, partitioning the data into just two regimes on slightly different criteria with respect to past growth than Tiao and Tsay.

Both the above papers showed that non-linear techniques were decisively superior to linear autoregressive representations of the data in terms of in-sample fit to post-war quarterly data on US GNP growth. However, the variance of the model error is barely less than the variance of the data, the former being 90 per cent of the latter in the Tiao and Tsay model and 88 per cent in Potter's best model. And this is the weakest possible test of forecasts, given that the fitted values of the model represent one-step ahead in-sample predictions.

With Michael Campbell (1997), I applied a non-parametric non-linear technique, namely a form of kernel smoothing, to the same data series. The ratio of the mean squared error to the variance of the data was 90 per cent for quarterly changes in US GNP, effectively identical to the two papers quoted above. We also built models for two, four and eight quarter differences in the log of the data, and carried out direct 2-, 4- and 8-step predictions, respectively. In other words, in predicting the 4-quarter growth rate from time t to time $t+4$, for example, only information up to time t was used, as it would be in a genuine forecasting situation. The results for the 2-step ahead direct predictions were very similar to those of the one-step, but the 4- and 8-step ones were even worse.

Both the actual forecasting record of macro models over a twenty-year period and the recent results from non-linear representations of the data suggest that most macroeconomic data series are inherently unpredictable. Campbell and Ormerod (1997) argue that this is an intrinsic property of the data, and one which is consistent with the Fisher hypothesis described above. In the past decade or so, important advances have been developed in non-linear signal processing, which it is perhaps helpful to think of as trying to identify the ratio of signal to noise in a data series. The approach is not concerned with the estimation of any particular model, but with the more general question of the signal-to-noise ratio. A very high level of noise relative to signal, for example, would suggest that in practice the data series would be indistinguishable from a random data series and no meaningful forecasts could be carried out. If a signal of sufficient strength exists, the practical problems of finding an appropriate representation of the data in a model may still be formidable, but at least in principle the effort can succeed.

A clear summary of the technique, known variously as singular value decomposition or singular spectrum analysis, is given in Mullin (1993) and a much more formal mathematical exposition is provided by Broomhead and King (1986). Other techniques in the physical sciences which are concerned with the existence of deterministic structure, such as the calculation of correlation dimension or of Lyapunov exponents, require far more data points than are available in macroeconomic data series. But singular value decomposition can be applied to noisy data series with lengths typically encountered in macroeconomics.

A brief overview of the technique is as follows. The essential concept is to form a delay matrix from a single time series in the following way. Given a series x_t, where t runs from 1 to n, choose a maximum delay d, and the first row of the matrix is $x_1, x_2,..., x_d$. The second row consists of $x_2, x_3,..., x_{d+1}$, and so on. Restricting ourselves for the purposes of illustration to three dimensions and avoiding formal mathematics, we can see that such a matrix contains rows (x_t, x_{t-1} and x_{t-2}). If we connect these points in sequence on a graph, the underlying structure of the data series may be revealed. For example, the attractors arising from the Lorenz equations can be reconstructed in this way.

The covariance matrix of the delay matrix is formed, and the eigenvalues of this matrix are calculated. The eigenvalues can be thought of heuristically as measuring the strength of movement of the series in the direction of the corresponding eigenvector. For example, if there were two eigenvalues which were large relative to the rest, this would indicate that a two-dimensional model might well give a good account of the series, since the series largely occupies, or fills out in, two directions only. The square roots of the eigenvalues are described as the singular spectrum. Noise which is present in the data series appears in the singular spectrum as a floor at the upper end. The significant values appear above the noise floor, and their number gives an upper limit on the embedding dimension of any attractor which might exist in the data.[1]

Applying this technique to the quarterly growth in real US GNP, for example, reveals a singular spectrum of the data which is similar, though not completely identical, to that of a random data series. There are no values which stand out clearly above the rest, nor do the values decline to a clear floor.[2] Similar results, which are even closer to those of random data series, are obtained by Campbell and Ormerod for annual growth in real GDP/GNP data for a range of OECD economies over a variety of sample periods.[3]

The interpretation of these findings is entirely consistent with the Fisher hypothesis. In practice, GNP data is indistinguishable in terms of the ratio of its signal to noise content from that of a random data series. There are simply too many factors which determine business cycles to enable forecasts to be made in which the variance of the error is significantly lower than that of the data.

It is very important to note that this does not necessarily mean that structural models of individual series cannot be built, using the word 'structural' in the following sense. To take a hypothetical example, suppose that fluctuations in GNP were determined solely by changes in the stock market index, with an error term whose variance was very small relative to that of the data. By construction a very good structural model of GNP could be built. But, as a vast literature shows, the stock market itself is unpredictable. So even though changes in GNP could be accounted for *ex post* by movements in the stock market, meaningful *ex ante* forecasts of GNP could not be carried out.[4]

The inherent unpredictability of movements in GNP undermines the concept of short-term, countercyclical policy. Criticisms of such policy are by no means new. The Lucas critique is well known, and many academics saw it at the time as dealing a fundamental blow to the use of conventional macro models in policy analysis. However, econometricians have rallied and have proposed ingenious ways of dealing with the criticism (see Clements and Hendry, 1995, for a summary of the key points).

Friedman argued against countercyclical policy many years ago on less esoteric grounds, invoking the simple formula for the variance of the sum of two series. If one series is GNP in the absence of countercyclical policy and the other is the effects of such policy, the variance of the two combined will only be less than that of GNP in the absence of policy if the covariance between the two is negative. Friedman was sceptical of policy makers having sufficient understanding of the economy to achieve such a result. Results obtained using modern non-linear signal processing techniques suggest that his intuition was correct, for if changes in GNP cannot be predicted successfully, it is only by chance that interventions designed to smooth out such fluctuations will succeed.

This latter point holds true even if good structural models of the economy can be built. Suppose, for example, that the conventional macro models were eventually able to agree on the impact on GNP of, say, a given change in government spending. This is far from being the case in reality, and even if it were, the results would not necessarily be true purely as a result of this hypothetical agreement. But suppose as well that the evidence amassed in favour of the result were impressive. Even in these circumstances it would not be possible to carry out short-term countercyclical policy which was consistently successful. For the unpredictable nature of the GNP data would mean that policy makers would often make the wrong kind of intervention, sometimes expanding or contracting when the economy would have moved the same way without intervention.

IV STRUCTURAL RELATIONSHIPS IN ECONOMICS

The distinction is made frequently in applied macroeconomics between forecasting and policy analysis. Clements and Hendry (1995) go so far as to suggest that separate models should be used for the two tasks. This is an idea with which modellers in the physical sciences are familiar.

The Santa Fe Institute time series competition presented researchers with a number of unidentified, highly non-linear time series of data. From the outset, three distinct goals were specified, as Gershenfeld and Weigend (1993) point out in their introduction to a description of the competition: first, to forecast, or to 'accurately predict the short-term evolution of the system'; second, to model or to 'find a description that accurately captures features of the long-term behaviour of the system'. (a clear distinction was made between these two aims, and indeed was demonstrated in the results of the competition); the third goal, described as 'system characterization', is much less familiar to economists but is the purpose, for example, of the signal processing technique described above. The aim is to determine fundamental properties of the system, such as the degree of randomness, which obviously overlaps with the aim of forecasting.

Conventional macroeconomic modellers increasingly make the distinction between forecasting and policy, but it is one which they have been compelled to make in the light of their actual forecasting performance. The econometric methodology of cointegration which underlies many of these models does, of course, place great emphasis on discovering long-run relationships. Each of the models used in the exercise of Bray *et al.* (1995), for example, is replete with such equations yet, as noted in section II above, despite over twenty years of intensive research effort, different models still give quite different results for the effects of various standard fiscal and monetary packages on the economy.

One possible explanation for the failure to agree on the effects of policies is that the models are overparameterized to a serious extent. Chatfield (1995) argues that this is a general fault of a great deal of statistical modelling. Advances in computer technology have far outstripped developments in statistical theory. As a result, not only are most models fitted to time-series data overparameterized, but the data-mining process by which they are selected means that there is no adequate statistical theory with which to judge the published models. Forty years ago, the physical process of computing even a simple regression was hard. Now, a graduate student, aided by menu-driven software, can generate a hundred versions of an equation in a single day.

This criticism applies with particular force to the data-driven approach which has dominated time-series econometric modelling in the UK for twenty years. The battery of tests which must be applied to equations before they are deemed fit to publish in respectable journals looks impressive, but the equations which are reported are invariably the result of intensive trawling of the data. The quoted test statistics can be thought of as design criteria. Models cannot be published unless they pass the currently fashionable list of tests. But the fact that the data are mined in order to produce a model which satisfies such tests undermines the whole value of the test procedure.

An important way in which this is reflected is in the parameter stability and forecasting performance of such equations once they confront genuine out-of-sample data – in other words, data which have not been used as part of the overall data sample with which apparent tests of stability have been constructed. Even inserting correct values of all the explanatory factors, the standard error of the model with genuine out-of-sample data exceeds that of the in-sample fit, both rapidly and substantially in many cases. This phenomenon is very well known amongst the practical model-building fraternity – purer academics are not often obliged to revisit the scene of their crimes once their article is published – but is rarely articulated in the public domain.

Overfitting can often be seen within the data sample used for estimation by applying simple techniques such as 'leave-one-out' regression. The 'leave-one-out' standard error of a regression is calculated as follows. If the equation is fitted on a sample of 1 to n data points, calculate the error for period 1 by fitting the equation on the sample 2 to n and backcast the first data point; for period 2, fit the model using period 1, and 3 to n, and so on. The resulting standard error is often distinctly larger than the one obtained fitting the model from 1 to n only. (The resulting residuals are often known as 'studentized' or 'deletion' residuals; see, for example, Atkinson, 1985.)

Another reflection of overfitting is the way in which published equations for, say, the consumption function change over time. A paper is published which heralds the 'truth' – the discovery of 'the' structural model which determines consumption. But a constant stream of such discoveries appears to be needed.

Despite these problems, conventional modelling proceeds on the basis of assumptions which implicitly rule out structural change, or regime shifts, in the economy. Mizon (1993), for example, describes the history of the

'LSE' econometric methodology, of which Hendry has been the foremost exponent, and the importance attached to parameter constancy over time. Interestingly, Hendry appears to have moved sharply way from this position very recently (see, for example, Clements and Hendry, 1995). He now argues that the assumption of parameter constancy is not appropriate, and that regime shifts are 'an important feature of the actual economy'.

A wider recognition of this fundamental feature of capitalist economies by what might be termed the conventional modelling community would represent a very important step forward in the research agenda. Ormerod (1994a), for example, examines Phillips curves in a wide range of OECD economies over the post-war period from the early 1950s to the early 1990s. For around 90 per cent of the time, very simple relationships between inflation and unemployment give a good account of behaviour. But the relationships are characterized by a small number of major shifts, which can be very abrupt in the sense that an economy can move from one apparently settled Phillips curve to another without a transition period. Mizon (1995), using quarterly data for the UK since 1966 to build a wage–price system, obtains a similar result in terms of the stability of his estimated relationships.

More generally, Ormerod (1994b, 1995) argues that, in local time neighbourhoods, straightforward linear models estimated with a very small number of parameters offer a good explanation of key macroeconomic relationships in the OECD economies. As well as the inflation/unemployment relationship, movements in unemployment itself, for example, can be accounted for by a simple relationship between unemployment and GDP growth (current and lagged) and lagged unemployment.

But such models almost invariably offer very poor explanations over longer time periods. Both these relationships have experienced a number of major shifts during the post-war period. There is a certain amount of common ground in the timing of these shifts, but the historical experience of each country is important in understanding both the timing and magnitude of the shifts. A small number of shocks in the post-war period, with the exact number varying from country to country, have caused permanent shifts in the inflation/unemployment and unemployment/growth relationships. But, by implication, the vast majority of the numerous shocks which affect on the economic system have had no such impact. The fashionable concept of hysteresis implies that every shock has a permanent effect, which is by no means the case.

A key task of any applied work should be to screen the data before any estimation is done at all, in order to identify periods when stable relationships might exist. Over such periods, very simple equations should be estimated which have a good theoretical basis, and with an absolute minimum of data mining. The role of statistical estimation is to parameterize such relationships, and to test the appropriateness of the shifts identified by prior screening of the data.

But the main task is to understand how and why shifts in such relationships take place, which requires careful analysis of the transition periods. It is here that a multidisciplinary approach becomes essential, and we move to questions which it is outside the power of time-series econometrics alone to resolve. And it is here that potentially the greatest policy gains can be made. An understanding, for example, of the circumstances in which the Phillips curve might shift would transform the potential power of governments.

NOTES

1. The eigenvectors associated with the dominant eigenvalues form the coordinate system onto which the data can be projected optimally.
2. This result is robust with respect to the choice of the delay factor, d. Even setting $d=50$ fails to reveal any significant values. The potential criticism of the approach by Vautard and Ghil (1989), namely that as d increases the number of significant eigenvalues may increase, does not apply to its use on macroeconomic data series, for significant values do not appear to exist at all.
3. These results are confirmed by Ormerod and Mounfield (2000) who apply random matrix theory, developed for the study of complex nuclear spectra, to annual GDP data for 17 OECD countries over the 1871–1994 period.
4. More generally, the use of multivariate rather than univariate techniques for forecasting does not of itself overcome the problem of a high noise-to-signal ratio in a data series. Such a ratio could arise for a number of reasons. The data could be genuinely random (for example, rolls of true dice); they could be determined by factors whose numbers are large relative to the available observations; or they could be determined by a small number of factors which themselves have high noise to signal ratios.