Twenty Years After: Econometrics, 1966-1986

Pagan, Adrian

Working Paper No. 94
August 1987

Twenty Years after:
Econometrics

Adrian Pagan

Rochester Center for Economic Research
Working Paper No. 94

TWENTY YEARS AFTER: ECONOMETRICS, 1966-1986*

by

Adrian Pagan

August 1987

Working Paper No. 94

*Paper presented at CORE's 20th Anniversary Conference, Louvain-La-Neuve, January 1987.

1. Setting the Scene

2. The Search for Standards

3. The Search for Consistency

3.1 Expectations
3.2 Consistency in Agents' Actions
3.3 Steady State Models
3.4 Achieving Steady State Consistency

4. The Searchfor Robustness

4.1 Making Inferences Robust to the Conditional Mean
4.2 Making Inferences Robust to the Second Moment
4.3 Robustness to Data Dependence
4.4 Robustness to Distributional Assumptions
4.5 Robustness to System Specification

5. The Search for the Individual

5.1 Models for the Individual
5.2 Selecting the Individual
5.3 The Duration of Events
5.4 From Micro to Macro

6. The Achievements: A Personal View

1. Setting the Scene

What has happened to econometrics in the twenty years since CORE was born? The external signs of growth are indeed impressive. In 1966 there were probably only three journals to be considered as vehicles for the regular publication of econometric methods - *Econometrica*, *Journal of the American Statistical Association*, and *The Review of Economics and Statistics*. By 1986 there were at least ten journals doing this, and even a specialist one in *Econometric Theory*. There is other evidence of growth. Johnston's text book, *Econometric Methods*, came out in 1963 and ran to some 300 pages. By contrast, the 1984 edition was 568 pages of much smaller print. The 1966 issues of *Econometrica* contained 910 pages, while the 1986 issues had 1503 pages. In 1966 it was probably the case that few members of the public would have even recognised the term econometrics; after the award of two Nobel Prizes in Economics to Tinbergen and Klein for their contributions to econometric modelling, public awareness is now much greater. So, on the external evidence, econometrics has flourished in the last twenty years.

A lot of this growth reflects the rapid advances in computing technology. It has always been true that the type of problems studied and work done in econometrics has been powerfully influenced by the available computation facilities. Techniques such as 2SLS and LIML were developed because it was too expensive to perform FIML; many methods for estimating equations with serial correlation had their origins in a desire to avoid the iterative procedures demanded by maximum likelihood. In 1966 there were no specialist econometric packages available for use; TSP was in the developmental stage, and those wishing to perform econometric analysis were forced to formulate their problems so as to employ the regression routines in packages such as OSIRIS. Thus both theory and practice were conditioned by the computational environment. Mostly what emerged up to 1970 was a second rather than first best choice.

All of this has changed dramatically in the last twenty years. Econometric packages have proliferated, and have become increasingly easy to use, particularly with the advent of micro-computers. Moreover, the modern package can estimate an enormous variety of models; some such as LIMDEP and SHAZAM even allowing users to customize the software. Consequently, the computational constraint has been significantly relaxed, rendering many "short-cuts" developed in the 1960's and 1970's merely of historical interest. Moreover, computational advances not only allowed the implementation of complex theoretical work, but also enabled solutions to problems too complex for theory - for example, many statistics such as the Durbin-Watson are ratios of quadratic forms in normal variables, and it was found possible to determine the numerical distribution of such quantities even if an analytic expression was intractable - Koerts and Abrahamse (1969), Durbin and Watson (1971).

This paper will not be directly concerned with the shift described above.

Important though it was, it is not primarily an intellectual tale. Furthermore, faced with the fact that the recently completed *Handbook of Econometrics* runs to 2093 pages, a good proportion of which deals with post-1966 research, some discussion about what is to be excluded from the purview of the paper is mandatory. My own interests naturally provide a first filter, but I have also adopted the principle of surveying those theoretical ideas which have had (or are likely to have) a significant impact upon the practice of econometrics. For this reason I have ignored developments in the fields of estimating markets in disequilibrium, continuous time models, and small sample theory. Excellent surveys of that material already exist in Quandt (1982), Bergstrom (1976) and Phillips (1982). It is my own opinion that, for one reason or another, research in these subjects has not yet had the impact upon practititioners which might have been anticipated in the early stages of their development.

Two other self-imposed guidelines operate. First, it is necessary to set the borderline between economic theory and econometrics. Many useful advances were made by the application of duality theory and calculus of variations to a study of business and consumer behaviour, and these were reflected in the types of specifications becoming the norm in applied econometric analysis. I have taken this literature as pertaining to economic theory, and have thereby limited my discussion to the estimation issues that arise. Secondly, one of the distinctive features of much of the work recounted below has been its origins in the mathematical statistics domain. Econometricians have translated these ideas into their own language and concerns. Sometimes the relationship has been symbiotic, with the interest displayed in the technique by our field re-awakening interest in it among mathematical statisticians. I have rarely tried to trace the intellectual heritage of the work described below, preferring to only locate the point of its introduction to econometrics.

With all these qualifications made, it is time to set out the structure of the paper. My title is borrowed from Alexandre Dumas' sequel to his classic "The Three Musketeers". That is a tale of the aging of the four musketeers, and the decline of their zest and vitality; it is a tale in which the events are precipitated by the flawed decisions of twenty years ago at the time of greatest triumph. Above all, it is a tale about a search to right the mistakes of the past. As such it constitutes a powerful metaphor for the changes that have beset econometrics in the past two decades. Twenty years ago, econometrics was probably in its golden age. Since then it has fragmented, particularly after the failure of econometric models in predicting developments following from the first oil price shock, and its spectacular growth is partly a function of the initiation of many searches aimed at rectifying what are now perceived as deficiencies inherent in its concerns at that time.

I have treated the focus of econometrics from 1966-1986 as four "searches" - for standards, consistency, robustness and the individual. It is interesting that, with the exception of the third, almost none of these was on the agenda of econometrics in 1966. That fact emhasises just how different econometrics has been in the last twenty years to the preceding twenty years. As Pesaran (1986) observes, the post-war developments in econometrics until the 1960's seem a very smooth ascending path, largely being a working out of the Cowles Commission programme in which careful attention was paid to the problems of implementation in the light of computational capabilities. By contrast, the last twenty years have witnessed tremendous diversity in econometric research. There is now no longer a single methodology for econometrics, and there is a much greater practical awareness of the limitations of any one. "Search" then seems an apt description of econometrics in the years we cover. The ultimate objective may be clear but the best route to get there is still to be found.

2. The Search for Standards

All modelling activity is based upon a set of conventions whose role it is to reduce a mass of potential outcomes down to a set of what are felt to be important ones. But, at times, these conventions may be seriously at variance with the data under analysis, and it becomes imperative to ascertain the validity of the conventions in any given context. It follows that normal practice should be for the reports on any modelling exercise to contain such discussion. In turn, this implies the need for a set of standards to evaluate the correspondence of conventions with outcomes. Unless such information is provided, any piece of research deserves to be treated with considerable scepticism.

The attempt to delineate such standards was a characteristic of the methodology for the modelling of economic time series advanced by Hendry (1980) and Hendry and Richard (1982). Their taxonomy is exhaustive, but I find it easier to discuss research on standards under five headings that correspond to the five main conventions underlying the general linear regression model.

$$y_t = x_t \beta + u_t . \tag{1}$$

(i) The conditional mean is linear in x_t only.

(ii) The variable x_t is uncorrelated with the error term u_t.

(iii) The conditional variance of the errors is constant.

(iv) The errors u_t exhibit no temporal dependence.

(v) (Sometimes) the errors u_t are normally distributed.

Clearly, there is no reason to suppose that *any* of these assumptions need be correct, and some measure of how accurate they are is desirable. To appreciate just how much of a change in attitude over this matter there has been in the last twenty years, it suffices to compare the paper by Huang in the April 1966 issue of *Econometrica* - a study of the U.S. housing market - with that by Engle et al (1987) twenty-one years later. In Huang's study almost no evidence is presented on the adequacy of the five assumptions above; each of his estimated equations is accompanied solely by an $\bar{R}^2$. By contrast, Engle et al incorporate an extensive range of diagnostic tests for failures in the crucial assumptions underlying their work. These two articles probably represent polar cases; some research in 1966 undoubtedly paid more attention to issues (i)-(v) than Huang did, while there are still studies in 1986 failing to be as responsive in meeting a set of standards as Engle et al are. Nevertheless, the profession has shifted in the latter direction; a comparison of the new editions of Johnston (1964), Kmenta (1986) and Kennedy (1985) with their previous versions makes this abundantly clear.

A summary of developments in this area might be done in a number of ways. One approach is to regard (i)-(v) above as statements about the assumed moments of the distribution of the dependent variable y_t; the degree of correspondence with the sample moments then provides the desired measures. Thus, if the conditional mean is postulated as $x_t\beta$, it is natural to ask if it might be augmented to $x_t\beta + z_t\gamma$ i.e. if $E(z_t'(y_t - x_t\beta))$ is equal to zero. The "sample" moment for this is $T^{-1} \sum_{t=1}^{T} z_t'(y_t - x_t\hat{\beta})$ where $\hat{\beta}$ is some suitable estimator of β, and this should be close to zero if the maintained model is to be acceptable. I have used this framework elsewhere - Pagan (1984) - terming it the "variable addition" strategy. Here I will adopt it to describe the major themes, and some extensions to the treatment will also be indicated.

(i) The Conditional Mean

Ramsey (1969) represents the most influential early contribution to an examination of this topic. He observed that neither the functional form nor the menu of variables appearing in the conditional mean could be asserted with much confidence. To provide some check on the maintained formats he proposed trying to add to the basic model, as z_t variables, polynomials in the regression predictions $\hat{y}_t$ (generally $\hat{y}^2_t$, $\hat{y}^3_t$ and $\hat{y}^4_t$). This became the RESET test. Subsequently, numerous variable addition tests appeared to assess the adequacy of particular aspects of the conditional mean specification. Each was derived by stipulating a plausible alternative model and then checking if the residuals of the maintained model exhibited any pattern that would be consistent with the alternative model being a better representation.

In 1978 Hausman advanced what appeared to be a quite distinct procedure. He advanced the idea that two estimators of the conditional mean parameters (β) be obtained, both being consistent if the maintained model was correct, but with different probability limits if it was incorrect, and that an index of model adequacy could then be based upon the difference in estimates. White (1980b) and Plosser et al. (1982) adopted this philosophy. In the former, OLS estimates of β were compared with weighted least squares values for β, where the weight given to the t'th observation was arbitrarily selected. In the second paper, the alternative estimate of β in the comparison came from regressing Δy_t against Δx_t, where Δ is the first difference operator. Both of these proposals are special cases of applying a known transformation matrix F to the matrix version of (1), $y = X\beta + u$, followed by a comparison of $\hat{\beta} = (X'X)^{-1}X'y$ and $\tilde{\beta} = (X'F'FX)^{-1}X'F'Fy$ (for White's case F is diagonal with the chosen weights as elements). Since $y = X\hat{\beta} + \hat{u}$, it follows that $\tilde{\beta} = \hat{\beta} + (X'F'FX)^{-1}X'F'F\hat{u}$ and $\tilde{\beta} - \hat{\beta} = (X'F'FX)^{-1}X'F'F\hat{u}$, making it a test based on $T^{-1} \Sigma z_t'\hat{u}_t = T^{-1} X'F'F\hat{u}$. As such, it is clear that these methods are oriented towards the detection of particular types of departures from the null hypothesis. Breusch and Godfrey (1986) have provided a comprehensive discussion of this.[1]

(ii) The nature of x_t

The second area of concern arises from the stochastic nature of x_t. A primary question is whether it is y_t or x_t which should be the dependent variable in any regression, or which variable is cause and which effect? Significant work has been done on this issue of "causality" since 1969. In that year Granger proposed an "empirical" definition that radically changed what had been an on-going and largely inconclusive debate ever since the Cowles Commission's reports. Granger proposed to define x_t as causing y_t if the ability to predict y_t using the past history of x_t and y_t together was greater than if only the past history of y_t had been exploited. In application this tended to be interpreted as whether the past history of x_t significantly augmented the regression of y_t upon its past values. Sims (1972) produced a modified version of this criterion by adding the requirement that no future values of x_t should have an influence as well.

To say that the idea was popular among applied econometricians would be a vast understatement; an enormous number of enquiries were conducted in this

[1]Hausman's principle is much more widely applicable than to just the linear regression model, and one of its most popular applications has been in checking the independence from irrelevant alternatives assumption in the multinominal logit model - Hausman and McFadden (1984).

framework during the 1970's on almost every conceivable economic time series. Much of it was very uncritical. As Leamer (1984) points out, the definition is more aptly one of the concept of "precedence", and the "causality" being investigated should be understood in that way. Otherwise, striking conclusions such as "Christmas card sales cause Christmas" (Kennedy (1985, p.64)) are all too likely. Moreover, many of these studies were conducted within a strictly bivariate framework, ignoring Granger's own warning about this: "If relevant data has not been included in this set, then spurious causality could arise" (p.429). A good proportion of those applying the method in a bivariate mode would seem to have totally forgotten one of the earliest maxims of economics statistics courses; ignored "third causes" can be the source of strong bivariate correlations. There was a lot of high-powered mathematics involved in the analysis of this topic, but I come away from a reading of it with the feeling that it was one of the most unfortunate turnings for econometrics in the last two decades, and that it has probably generated more nonsense results than anything else done during that time.

To many econometricians what was of interest in Granger's work was not "causality" *per se*, but the idea that it could be used to determine if x_t was "exogenous". The "new classical economists" frequently asserted that x_t was "econometrically exogenous" if and only if x_t was not Granger-caused by y_t i.e. x_t was a function only of its own-history (see Sargent (1976, p.544) for example), and berated econometricians for not doing such checks - Lucas and Sargent (1981, p.302). Yet, on almost any reasonable definition of "econometrically exogenous", it was difficult to understand why Granger causality was required. For example, if $x_t = \alpha x_{t-1} + \gamma y_{t-1} + v_t$, where v_t is uncorrelated with u_t, it is not necessary for γ to be zero to ensure that x_t is uncorrelated with u_t. This stance by Sargent, Sims et al. has always puzzled me, and the only way I have made sense of it is to argue that adherents to this position wish to make *no* assumptions about the degree of serial correlation in the error terms. But, I must admit that, if this is what Sims (1972) intended in his original discussion of the problem, it is not at all clear.

Confusion over questions of "exogeneity" reigned in applied work during the 1970's, and one could almost sense a sigh of relief when Engle et al (1983) appeared, with a simple summary of the main "exogeneity" concepts in the literature. By far the most important theme of the paper was to take the joint density for (y_t, x_t), conditional upon all past data, $f_{t-1}(y_t, x_t; \lambda)$, and ask if it could be factored as $f_{t-1}(y_t / x_t; \lambda_1)\ f_{t-1}(x_t; \lambda_2)$ i.e. do the parameters entering in the conditional density also appear in the marginal density for x_t? If not, they termed such a situation one in which x_t was "weakly exogenous" for λ_1; a proposal wholly compatible with the original Cowles Commission idea of exogeneity, as exemplified by Koopmans (1950). Furthermore, if $f_{t-1}(x_t;\lambda_2) = f^*_{t-1}(x_t; \lambda_2)$, where the conditioning in $f^*_{t-1}(.)$ is now upon the past history of x_t alone, Granger causality or "strong exogeneity" prevails. Because of the nested nature of this decomposition, it is immediately apparent that Granger causality is too strong for efficient estimation of λ (or any functions of it).

Still, the question remains, what is an appropriate test for weak exogeneity? A standard reply in many articles has been the presentation of the Wu-Hausman statistic (proposed informally by Durbin (1954)). Wu (1973) suggested that a comparison be made between the OLS estimator $\hat{\beta}$ and an instrumental variables estimator $\tilde{\beta} = (Z'X)^{-1}Z'y$. Once again, after substituting $y=X\hat{\beta}+\hat{u}$ into the formula for $\tilde{\beta}$, $\tilde{\beta} - \hat{\beta} = (Z'X)^{-1}Z'\hat{u}$, and therefore the test involves the sample correlation of the instruments with OLS residuals assuming x_t is weakly exogenous. Another interpretation of Wu's proposal is to add Z to the original model, and this is what Hausman (1978) did. From the nature of the sample moment under test, it is apparent that what is being examined is the validity of the instruments, and that in turn is dependent upon the validity of the specification of the marginal density $f_{t-1}(x_t,\lambda_2)$.

(iii) Conditional Error Variances

Investigation into the failure of higher order moment assumptions for the errors was popular, although what progress there was in this area largely involved the re-interpretation of already existing tests, rather than the development of new ones. Thus, the design of diagnostic techniques for heteroskedasticity in Breusch and Pagan (1979) and Godfrey (1978b) had antecedents in the literature; the novel element supplied by these papers was the connection with Lagrange Multiplier (LM) statistics. All of the tests that were proposed here essentially involved checking if the correlation of some nominated variables z_t with the centered squared errors was zero i.e. was $T^{-1}\Sigma z_t'[(y_t - x_t\beta)^2 - \sigma^2]$ close to zero? In the LM versions, $\theta' = (\beta'\sigma^2)$ were replaced by the OLS estimators $\hat{\theta}$ but, as observed in Pagan (1986), any root -T consistent estimator of θ could be used without affecting the asymptotic power properties of the test, because the limiting distribution of the statistic above does not depend upon the limiting distribution of the θ estimators. However, in small samples the choice may be important, and Evans and King (1985) proposed estimating θ from a weighted regression.

Perhaps the most interesting development under this heading was the recognition that the variance of u_t might be unconditionally constant, but nevertheless vary conditionally, e.g. the conditional variance $E_{t-1}(u_t^2) = \sigma^2 + \alpha y_{t-1}^2$ has $E(u_t^2)$ as a constant whenever y_t is a covariance stationary process. Engle (1982) popularised this distinction, naming the conditional case Autoregressive Conditional Heteroskedasticity (ARCH). Lately, it has become common to compute tests for certain variants of this class of models, particularly those that have $E_{t-1}(u_t^2) = \sigma^2 + \sum_{j=1}^{M} \alpha_j(y_{t-j} - x_{t-j}\beta)^2$; the idea here being that large errors in the past can be expected to create greater uncertainty in present decisions

(iv) Temporal Dependence

Unlike the analysis of the second moment of u_t, work on temporal dependence yielded many new insights. In 1966, the principal way of assessing temporal dependence was with the Durbin-Watson statistic, but it was already known that there were difficulties in applying it to dynamic models in which lagged values of the regressand appeared. The first correct solution to the problem was presented in Durbin (1970). He observed that the Durbin-Watson statistic can be thought of as based upon the sample moment $T^{-1} \Sigma \hat{u}_t \hat{u}_{t-1}$, and so its limiting distribution may well depend upon that of $T^{1/2}(\hat{\beta} - \beta)$. Indeed, this is exactly what happens when x_t contains y_{t-1}, and Durbin's solution was to make an allowance for the fact that β is estimated rather than known.

An appreciation of this principle cleared the way for an extension of the traditional tests for serial correlation to other contexts e.g. when $\hat{\beta}$ was an instrumental variable estimator as in Godfrey (1978a). Of course, when the limiting distribution is independent of $T^{1/2}(\hat{\beta} - \beta)$, other choices of $\hat{\beta}$ than the OLS one may be desirable for good small sample performance. Berenblut and Webb (1973) suggested that $\hat{\beta}$ be replaced by $\tilde{\beta}$ formed by regressing Δy_t against Δx_t, while King (1985) adopts an estimate found by regressing $y_t - \lambda y_{t-1}$ against $x_t - \lambda x_{t-1}$, where λ is a pre-specified value. He demonstrates that the small sample performance of such a modified statistic is generally superior to that for the Durbin-Watson test.

The other major extension in this area arose from the steady movement of empirical work from yearly to quarterly and even monthly data. Quarterly data meant that variables were likely to exhibit a fourth order serial correlation pattern, and this makes an examination for only first order serial correlation inadequate. Ideally, it is desirable to look at the complete autocorrelation function of the residuals, but specialised variants were developed e.g. for fourth order by Wallis (1972).[2]

(v) Distribution of the Errors

Normality of the errors has traditionally been queried by examining the coefficients of kurtosis and skewness of the residuals. The distribution of these depends upon that of $\hat{\beta}$, and it was Bowman and Shenton (1975) and Bera and Jarque (1981) who provided the requisite adjustments; the latter being particularly interesting in its derivation of the test statistics via the LM principle applied to alternative distributions falling within the Pearson family. As checks on *normality*, these indices are not entirely satisfactory, owing to the existence of distributions, notably Tukey's λ distribution (Joiner and Rosenblatt (1971)), which exhibit identical skewness and kurtosis to the normal but which differ in the tails. For regression models this difference is not of great import, but for models in which the dependent variable is truncated or censored - for example in the Tobit model where only values of y_t corresponding to $x_t\beta + u_t \geq 0$ are observed - it is the shape of the whole distribution which is important. Because of this fact comprehensive tests of normality have appeared in the literature - Heckman (1984) advocated

[2]Although little used, the investigation of non-linear patterns in residuals may be important in certain contexts. Thus evidence on the "efficient markets hypothesis" should not be based solely on the detection of *linear* patterns of autocorrelation; the hypothesis rules out non-linear ones as well. I found evidence of a non-linear weekly effect in stock market data in Pagan (1978) and Hinich and Patterson (1985) reach a similar conclusion.

the chi-square test comparing actual and predicted fractiles of empirical distribution functions, and Andrews (1985) generalised this in a number of directions, notably to allow the selection of fractiles to be sample-based. Tauchen (1985) has a similar proposal, based on a discrete number of fractiles.

A feature of the 1970's, treated in a later section, was the movement away from the standard regression model to situations involving latent variables or categorical data. Hence, there is frequently no continuously measured dependent variable, and it becomes imperative that indices of adequacy be devised from a perspective that is not limited by regression concepts. The most important papers concerned with this task were Tauchen (1985) and Newey (1985). These authors argued that, if there exists a vector of functions $m(w_t,\theta)$, (θ being the parameters of the maintained model and w_t the data) with the property that $E(m(w_t,\theta)) = 0$ when the maintained model is valid, then $\hat{\tau} = T^{-1}\Sigma m(w_t,\hat{\theta})$ is a suitable vector of indices to test the adequacy of the original model. As illustrations, $m(w_t,\ \theta) = z_t'(y_t - x_t\ \beta)$ would be the choice for specification error in the mean of the regression model, while $m(w_t,\theta) \ = \ z'_t[(y_t - x_t\beta)^2 - \sigma^2]$ would provide tests for heteroskedasticity in it.

What is the appeal of this approach? Many econometric estimators of θ satisfy a set of equations $T^{-1}\Sigma d(w_t,\hat{\theta}) = 0$; in the case of maximum likelihood estimation the d(.) are the scores but, as discussed later, instrumental variable estimators also have this format. Consequently, if $\phi(w,\theta,\tau) = T^{-1}\Sigma[d(w_t,\theta) \mid (m(w_t,\theta) - \tau)]$, the estimates of θ and τ found by setting $\phi(w,\theta,\tau) = 0$ are just $\hat{\theta}$ and $\hat{\tau} = T^{-1}\Sigma m(w_t,\hat{\theta})$, and $\hat{\tau}$ may be treated as a *parameter* in an extended estimation problem. If computer programs are available to solve $T^{-1}\Sigma d(w_t,\theta) = 0$, immediate application to the task of solving $\phi(w,\hat{\theta},\hat{\tau}) = 0$ is possible, and standard tests of $\tau = 0$ may be performed. Actually, when $\hat{\theta}$ is an MLE, after some suitable partitioning and use of asymptotic theory, both Tauchen and Newey give a simpler test procedure for $\tau = 0$; regress $m(w_t, \hat{\theta})$ against unity and $d(w_t,\hat{\theta})$, checking if the intercepts in these regressions are zero.[3] If $\hat{\theta}$ is not an MLE, Newey provides an expression for the limiting distribution of $\hat{\tau}$.

By adopting the Newey/Tauchen formulation it is easy to derive indices of inadequacy for cases other than the linear regression model. All that need be done is to select a suitable m(.) function. To date the main extensions have been to the situation where the general linear model is the underlying structure but output from it is censored. The dependent variable in (1) becomes a latent variable y_t^* and the actual observations y_t do not equal y_t^* for part of the sample; in the Tobit model only non-zero y_t^* are available, while in the Probit model y_t is binary.

Suppose it is possible to propose an alternative model differing from the

---

[3]Both authors assume that $d(w_t,\theta)$ and $m(w_t,\theta)$ are independently distributed random variables. However, it is only necessary that such quantities be martingale differences.

maintained version by the presence of extra parameters ψ. The theory of the E-M algorithm (Dempster et al. (1977)) contains the useful result that the actual score with respect to ψ is given by the conditional expectation of the score appropriate to the latent variable model. That is, if $L*$ is the likelihood of the latent data and L that of the observed data, $\partial \log L / \partial\psi = E(\partial \log L* / \partial\psi | \text{obs data})$ (Chow (1983) has a good discussion of this). Consequently, a suitable m(.) would be this conditional expectation, and it forms the basis of diagnostic tests given in Bera et al. (1984), Lee and Maddala (1985) and Gourieroux et al. (1985).

One interesting aspect of these developments has been the definition of "generalised residuals", in the sense of Cox and Snell (1968). To get these, consider augmenting (1) with z_t when y_t^* is latent and u_t is n.i.d. (0,1). From the principle connecting the observed and unobserved scores, $\partial \log L/\partial\psi = E[\Sigma z_t'(y_t^* - x_t\beta) \mid \text{obs data}] = \Sigma z_t' E[(y_t^* - x_t\beta) \mid \text{obs data})]$. In the linear regression case the corresponding score would be $\Sigma z_t'(y_t - x_t\beta)$, and it is fitting that $E[(y_t^* - x_t\beta) \mid \text{obs data})]$, evaluated at $\hat{\beta}$, be regarded as a *generalised* residual. For the probit model $E[(y_t^* - x_t\beta) \mid \text{obs data}] = y_t F(x_t\beta)^{-1} f(x_t\beta) + (1 - y_t)F(-x_t\beta)f(x_t\beta)$, where F(.) and $f(.)$ are the cumulative normal distribution and standard normal density evaluated at ".", and this quantity forms the basis of the diagnostic tests set out in Davidson and MacKinnon (1984). Once the generalised residuals are computed they may be used for checking the assumptions of homoskedasticity, normality etc. in much the same way as the ordinary residuals were used earlier - Chesher and Irish (1984).[4] Whether it is worth specifically obtaining the generalised

[4]Since $E((y_t^* - x_t\beta) | \text{obs data}) = E(u_t | \text{obs data})$, the moments of u_t can be found as in Amemiya (1973), and it is these which are employed.

residuals is a moot point. Unlike the ordinary residuals it is hard to interpret plots of them, and so their utility is mainly that of a pedagogic device for explaining the intuition behind diagnostic tests constructed with m(.) as the conditional scores.

Most of the work described above implicitly accords to the maintained hypothesis a special status. The role of alternatives is to expose inadequacies in the current formulation, and they are rarely regarded as being worthy of estimation or reporting. For this reason, the alternatives typically reflect statistical rather than economic theory, although it is questionable if the process of econometric modelling is capable of being characterised in this way. Frequently, economic reasoning will supply a number of potential explanations, with no single one capable of being designated as the maintained variant. A method of directly comparing such models is therefore clearly desirable.

This consideration gave rise to a good deal of interest in the question of comparing non-nested models. Dhrymes et al. (1972) mentioned work by Cox (1961), (1962) when looking at methods for evaluating models, but it was not until after Pesaran's (1974) paper on the choice between two non-nested linear regression models that applications emerged. Thereafter, a voluminous literature developed, surveyed in McAleer and Pesaran (1986). Most of it may be treated within the same framework as that utilized previously in the construction of diagnostic tests. It is now necessary though to partition θ into θ_1 and θ_2, corresponding to the two models entering the pair-wise comparisons. Auxiliary functions are still appended to the set of equations $T^{-1}\Sigma d(w_t,\hat{\theta}) = 0$, and a test statistic based on $\hat{\tau}$ constructed. What complicates the analysis is that $E(d(w_t,\theta)) = 0$ for only a subset of equations, as one of the models is taken to be invalid. Consequently, if $T^{1/2}(\hat{\theta}_1 - \theta_1)$ is $\mathcal{N}(o,v_{11})$, $T^{1/2}(\hat{\theta}_2 - \theta_2^*)$ will be $\mathcal{N}(o,v_{22})$, where θ_2^* is referred to as the "pseudo-true value" of θ_2 (Sawa (1978)) and $\Sigma E(d(w_t, \theta_1, \theta_2^*)) = 0$. Derivation of the limiting distributions of τ therefore need to be done carefully but, when $d(w_t, \theta)$ are the scores of each likelihood, analyses of the limiting distribution of $T^{1/2}(\theta_2 - \theta_2^*)$ have been introduced into the econometric literature by White (1982a) and Gourieroux et al. (1984). This theory was applied to $\hat{\tau}$ in Gourieroux et al. (1983) and Mizon and Richard (1986).

What differentiates the contributions to this literature is not the choice of $d(w_t,\theta)$ but the specification of $T^{-1}\Sigma m(w_t,\theta)$.[5] For simplicity, consider the choice between two linear regression models $y_t = x_t\beta + u_{1t}$ and $y_t = z_t\gamma + u_{2t}$. Various selections for $m(w_t,\theta)$ are possible.[6] First, take $m(.) = z_t'(y_t - x_t\beta)$, since the error term $(y_t - x_t\beta)$ should be uncorrelated with any regressor in z_t not in x_t. This yields the F-test for non-nested models. Alternatively, set $m(.)$ to $(z_t\gamma)(y_t - x_t\beta)$, a single auxiliary function, to get the J-test of Davidson and MacKinnon (1981) - frequently implemented by

[5]An exception is Singleton (1985) who looked at $\hat{\theta}$ coming from the GMM estimators discussed later.

[6]In what follows comparisons are done with $y_t = x_t\beta + u_{1t}$ as the maintained model, but the process needs to be reversed as well since there is no reason to give it any special status.

adding $z_t\hat{\gamma}$ to $y_t = x_t\beta + u_{1t}$ and then testing for its significance.[7] Another choice would be $\Sigma m(w_t,\theta) = L(\gamma) - E_\beta L(\gamma)$, where $L(\gamma)$ is the log likelihood of $y_t = z_t\gamma + u_{2t}$ while $E_\beta L(\gamma)$ is its expectation under the first model. This is the basis of Cox's statistic. Mizon (1984) and Mizon and Richard (1986) have noted that there are many ways to formulate suitable m(.) functions, based upon scores, likelihoods, parameter estimates etc., and each of these corresponds to asking the question of whether one model encompasses the other with respect to the particular m(.) function adopted.

There is no doubt in my mind that this work on adequacy was vital to improving the standard of applied econometrics. In some respects its role is negative; having a model designed to meet the criteria doesn't mean it is a good model, but having a model that fails them is indicative of a bad one. A major deficiency with econometric research of the 1960's, well exemplified by Huang's paper, was the impossibility of inferring from any reported material whether the model was an adequate representation of the data. Over the years I have been continually impressed by the role of these tests in ferretting out poor models, to such an extent that I now feel very uncomfortable with conclusions drawn on the basis of research which has been subject to little critical assessment through them.

---

[7]Adjustments to the J-test, aimed at getting a closer correspondence of asymptotic theory with small sample behaviour can be found in Godfrey and Pesaran (1983). In Pagan (1984) I suggested a way of finding the Type I error of the J-test by Monte Carlo methods, but James MacKinnon has pointed out to me that the argument given there is incorrect.

3. The Search for Consistency

As mentioned in the introduction, econometric research in the 1970's was frequently stimulated by the perceived failure of large-scale econometric models to predict well after the first oil price shock. One response to this phenomenon would be to claim that such an outcome was inevitable, either because the models had been constructed with a flawed methodology or because very little attention had been paid to whether they met acceptable standards. I certainly believe that there is more than a grain of truth in this judgement. But neither point is specific to large-scale macro models, and it is therefore not surprising that some of the most active research areas in econometrics dealt directly with what were perceived to be inherent weaknesses in the large-scale models themselves. What motivated this research was the desire to make those models internally consistent, with the concomitant belief that re-structuring along the new lines would lead to better performance. In making these models consistent, a range of econometric issues arose, some of which were solved and some of which still constitute on-going research programs. This section distinguishes a number of ways in which internal consistency was sought.

3.1 Expectations

The need for consistency first arose from a consideration of the way expectations were normally modelled. Standard empirical macro-economic models based output decisions upon the ratio of the wage rate to actual prices while, following the modification of the Phillips curve introduced by Friedman and Phelps, the wage rate was determined by price expectations. Suppose these price expectations are represented as a combination of past and present prices i.e. $p_t^e = \sum_{j=o}^{K} w_j p_{t-j}$ $(0 < K < \infty)$. Then any rise in p_t is only partly reflected in p_t^e, wages rise less than prices, and output expands. Workers are "fooled" into supplying more labour even though the actual real wage declines. With the operation of this "Keynesian" mechanism, fiscal policy is expansionary.

Although it is quite plausible that such "mistakes" are made, what was implausible was their persistence, as that can be construed as a defective learning process on the part of economic agents. Believers in "rational economic man" were decidedly uncomfortable with such an implication, and proposed that no model should embody the possibility of a consistent pattern of mistakes.

Borrowing Muth's (1961) phrase of "rational expectations", Lucas (1972) argued that expectations in a model should be rational and have the property that they be the predictions of the model for the variable about which expectations are being formed. To present the implications of this viewpoint and to discuss the literature surrounding it, consider the following simple demand and supply model for a commodity.

$$d_t = \alpha p_t + \beta i_t + e_t \tag{2a}$$

$$s_t = \gamma p_t^e + \delta w_t + \epsilon_t \tag{2b}$$

$$s_t = d_t = q_t \tag{3}$$

(2) and (3) had d_t, s_t and q_t as quantity demanded, supplied and transacted respectively, p_t and $p^e{}_t$ are actual and expected prices; i_t is real income; w_t the "weather"; e_t and ϵ_t are zero-mean white noise error terms. Older market models of commodities set $p_t^e = p_{t-1}$, but in the new view p_t^e was replaced by $E(p_t|i_t,w_t)$. To find $E(p_t|i_t,w_t)$, get the reduced form equation for p_t,

$$p_t = \pi_1 i_t + \pi_2 w_t + \pi_3 e_t + \pi_4 \epsilon_t, \tag{4}$$

and take the conditional expectation,

$$p_t^e = E(p_t|i_t,w_t) = \pi_1 i_t + \pi_2 w_t. \tag{5}$$

Notice that there are no systematic mistakes in price expectations in this model since $p_t - E(p_t|i_t,w_t) = \pi_3 e_t + \pi_4 \epsilon_t = v_t$, is a zero mean white noise.

Econometricians attacked the estimation problems caused by rational expectations (RE) with gusto. McCallum (1976) suggested that $p_t - v_t$ be substituted for $p^e{}_t$ in (2b), making it

$$s_t = \gamma p_t + \delta w_t + \eta_t, \tag{6}$$

and then estimate (2a), (3) and (6) jointly by two stage least squares. Others proposed Three Stage Least Squares and FIML - Wallis (1980), Wickens (1982).

Thus, the presence of RE's in macro-economic systems was effectively dealt with by a transformation replacing the expectation by an endogenous variable, followed by an application of a simultaneous equation estimator. But the simplicity of this solution raises a number of technical issues. First, is the new system identified? Although the number of endogenous variables may remain unchanged (p_t was frequently already an endogenous variable before, as it is in (2a)), the number of predetermined variables was generally smaller than when p_t was determined in an extrapolative way. For example, if $p^e{}_t = p_{t-1}$ in (2b) there would be three potential instruments $\{p_{t-1}$, w_t and $i_t\}$ and (2) is over-identified, but in the RE case there are only two, making (2) just identified. With RE's in a system, care must be exercised to ensure that there really are enough available instruments to enable estimation to proceed - Revankar (1980), Pesaran (1981).

Complications arise if (2b) features forward looking expectations in that it is p^e_{t+1} rather than p^e_t that drives supply decisions. Then η_t in (6) is composed of $\epsilon_t - \gamma v_{t+1}$ and has the autocorrelation function of an MA(1) process. Although the instrumental variable solution is still available, the MA-type error term should be accounted for in efficient estimation. However, if the composite error term η_t was just replaced by an MA(1), $\xi_t + \alpha\xi_{t-1}$, there is no guarantee that $E(\xi_t | i_t, w_t) = 0$ even though it is true for the individual components. Hence, the estimation problem is not a straightforward one and various suggestions were made to improve the efficiency of estimation. Pesaran (1987) and Wickens (1986) provide detailed treatments of this literature, emphasizing that there is a close connection between the estimation problem and the difficulties of finding unique solutions in forward-looking expectations models.

A second issue concerned the possibility of testing if the assumption of RE's was correct. Again, the simultaneous equation perspective provided an affirmative answer. To see why, begin with a single structural equation:

$$y_{1t} = y_{2t}\gamma + x_{12}\beta + e_t, \tag{7}$$

where y_{2t} are included endogenous variables, x_{1t} are included predetermined variables, and e_t is the disturbance term. The reduced form equations for y_{2t} are:

$$y_{2t} = x_{1t}\pi_{21} + x_{2t}\pi_{22} + v_{2t} , \tag{8}$$

enabling substitution for y_{2t} in (7) to give:

$$y_{1t} = x_{1t}(\pi_{21}\gamma + \beta) + x_{2t}\pi_{22}\gamma + v_{1t} \tag{9}$$

$$= x_{1t}\pi_{11} + x_{2t}\pi_{12} + v_{2t}, \tag{10}$$

the reduced form equations for y_{1t}. Clearly, the presence of y_{2t} on the right hand side of (7) induces the restrictions $\pi_{11} = \pi_{21}\gamma + \beta$ and $\pi_{22}\gamma = \pi_{12}$. Since RE's causes an endogenous variable to appear in an equation, it also imposes restrictions between the structural and reduced form parameters. By inserting OLS estimates of π_{ij}, and the 2SLS estimate of γ and β, it is possible to test if $\pi_{11} = \pi_{21}\gamma + \beta$ and $\pi_{22}\gamma = \pi_{12}$. Byron (1970) observed that this was a way of checking the validity of structural specifications, while Hoffman and Schmidt (1981) applied it to the RE situation.

To some extent I think that this approach is not as useful as it might be. Because it essentially compares the restricted and unrestricted reduced form parameters, it is important that the reduced form be general enough to allow for the plausible range of alternative hypotheses. For example, if the alternative in (2b) above is that $p_t^e = p_{t-1}$, the reduced form regressors should include p_{t-1}, whereas it is excluded under the RE formulation. A related difficulty resides in the fact that the choice to be made is really between RE's and some extrapolative mechanism, and this presumably should be done directly rather than via a reduced form. Technically, what is involved are two non-nested models, in that one cannot be found as the limit of the other, and it would be better to do comparisons utilizing the techniques outlined in the preceding section.

Perhaps the most important developments to emerge out of this interest in estimating models with RE's was the re-focussing of attention upon the role of economic theory in supplying a set of orthogonality relations between functions of an observed variable and an unknown error term. Sometimes, these relations come from an optimising framework - Kennan (1979). In the context of (2a) and (6), the orthogonality conditions are that $E(z_t\eta_t) = 0$ and $E(z_te_t) = 0$, where $z_t = (w_t \,\vdots\, i_t)$, since $\eta_t = \epsilon_t - \gamma v_t$ and, as w_t and i_t are both assumed to be in the agents' information set, $E(w_tv_t) = E(i_tv_t) = 0$. A prominent advantage of this approach is that not all of the information used in forming expectations need be specified - a sub-set of the orthogonality relations is all that is required for consistent estimation. Replacing η_t in (6) by $q_t - (p_t \; w_t) \begin{bmatrix} \gamma \\ \delta \end{bmatrix} = q_t - x_t\phi$, the orthogonality restrictions are $E(z_t'(q_t - x_t\phi)) = 0$. A sample analogue is $T^{-1}\Sigma z_t'(q_t - x_t\phi) = 0$. Solving this last equation for ϕ, assuming that $T^{-1}\Sigma z_t'x_t$ is non-zero, provides the instrumental variables estimator $\hat{\phi} = (\Sigma z_t'x_t)^{-1}\Sigma z_t'q_t$.

Interesting questions arise when there are more orthogonality conditions than parameters ϕ i.e. $\dim(z_t) > \dim(x_t)$. Sargan (1958) demonstrated that, within the class of consistent instrumental variable estimators of ϕ, the variance minimising estimator is formed from $z_t = \hat{x}_t$, the predictions from the regression of x_t against all possible instruments. Sargan termed this estimator the Generalised Instrumental Variables Estimator (GIVE). What was not fully grasped at the time, although quite clear from Sargan's proof, is that the optimality of GIVE depends crucially upon the assumption of the errors in the orthogonality conditions being i.i.d. $(0,\sigma^2)$ i.e. the errors should exhibit no temporal dependence and have a constant conditional variance. To appreciate the restrictiveness of that assumption, suppose p^e_{t+1} rather than p^e_t was in (2b). Then $E(v_{t+1} \mid w_t, i_t) = 0$, v_t may be a one-dependent process, and so in turn may η_t. Under such circumstances greater estimator efficiency is possible by combining the instruments in a way that recognises the temporal dependence in the errors. Cumby et al. (1983), White (1982b) and Hansen (1982) all gave the correct weighting scheme, with the latter describing the estimators as Generalised Method of Moments (GMM), thereby emphasising a connection with the most basic estimation method dealt with in statistics.

GMM was not a major advance over existing estimation methods, but it represented a successful re-interpretation of existing ones and was a concept of enormous power when it came to unifying a scattered literature. As an added bonus, the consistency and asymptotic normality of such estimators was established by White and Hansen for a number of types of variation in economic variables. Applications of the idea to as diverse a range of issues as sequential estimates (Newey (1984)), non-linear RE models (Hansen and Singleton (1982)) and multi-period probit models (Avery et al. (1983)) demonstrate this fact.

3.2 Consistency in Agents' Actions

Far too many models are built without a consistent framework for agents' actions. It is not uncommon to see separate equations in macro models for prices, output and labour demand, all deriving from different theoretical perspectives and with little recognition of the fact that these decisions are *interdependent*. Brainard and Tobin (1968) constituted a path-breaking paper setting out the implications of a failure to do so. They were concerned with portfolio allocation models in which the demand for each asset in a portfolio was related to the size of the portfolio and relative returns. Since the sum of all demands must in aggregate equal the size of the portfolio, Brainard and Tobin pointed out that it was impossible to independently specify each equation while retaining consistency. The demand for the n'th asset must be a combination of the demand functions for the other (n-1) assets. Consequently, unless care is exercised in specifying demands for the first (n-1) assets, the format of the derived demand for the n'th asset may be unacceptable.

Macro models exhibit similar problems, arising from the specifications of decisions taken by both the household and business sectors. For the latter, it is important to account for prices, output and employment. Without a consistent framework model performance becomes questionable. To illustrate, in the Australian model, NIF-10S, in some simulations three quarters of the output response does not require any increase in primary factor usage. With output determined rigidly by a production function, this would be implausible. Actually, in NIF-10S demands are derived from a production function, but the implied restrictions are never enforced. "Consistency" is forthcoming by the presence of a missing factor of production - dubbed "phonium" by Challen and Fitzgerald (1984) (for "off-production-function-onium").

A lack of consistency can show up in other ways. All too often, equations are formulated as relations of flow variables with other flows, ignoring the fact that corresponding to every flow there is a stock. Stock movements would be expected to feed-back to flow decisions, and this principle should be reflected in any chosen specification. Failure to do so invariably leads to poor model performance.

Ideally, consistency would be imposed by the derivation of estimable equations from a well-defined optimisation process. The movement towards

Euler equation methods for this task is a step in that direction. Unfortunately, the quality of economic data is rarely good enough to enable accurate calibration of such precise models. More than that, for the examination of issues, like the effects coming from a tariff change, the data may not even be available. To address issues of this nature "hybrid" models, designed to exhibit a high degree of internal consistency and calibrated partly by econometric models and partly by judgement, have become popular. Examples are the macro models of Rose and Selody (1985), Braynton and Mausskopf (1985), Murphy et al. (1986); micro-based versions encompass the computable general equilibrium models of Piggot and Whalley (1985) and the ORANI model of the Australian economy - Dixon et al. (1982).

I confess to a belief that hybrid modelling is an influential trend in econometrics. The development of micro-computers now makes it possible for a wide range of people to formulate their own model and to easily analyse issues using them. "Personal" rather than "public" models may become dominant in the next decade. Essentially this is a numerical variant of the sort of thing economists have done analytically for years: set up a theoretical model and then compute partial derivatives to assess the consequences of various actions. What the hybrid models add is the possibility of incorporating a richer set of environmental details at the cost of having to specify particular functional forms and parameter values. The role of econometrics in this case is to specify "best-bet" values for the parameters, thereby achieving some correspondence with actual outcomes. If this trend becomes dominant, there will arise the need to describe the sensitivity of model responses to assumed parameter values. This is a very interesting area of research, and largely an untapped one, although the question has been addressed by Kuh et al. (1985), Harrison (1985), Kalaba and Tesfatsion (1985) and Pagan and Shannon (1985).

3.3 Steady State Models

"Small is beautiful" became the password in macroeconometric modelling during the past decade. Reduction in dimensionality reflected a growing conviction that, whenever difficulties arose within a large model, it was very hard to isolate and understand the causes, whereas a small model might be analysed quite accurately with the aid of textbook macro models. Dissent was also based on deeper considerations, since it became increasingly obvious that the models deviated quite fundamentally from the consensus view of the properties a good model should possess. Many of these models appeared not to have any steady state solution; once shocked convergence back to any point or path rarely eventuated. Although this feature may not be important for forecasting applications of models, it is nevertheless disturbing to see a gap emerge between theory and practice.

The first serious attempts to bridge this gap were the models growing out of the International Money Project at the London School of Economics - Bergstrom (1976) and Jonson et al. (1977). Basic to their design was the selection of specifications ensuring convergence back to a pre-specified growth path after any policy action. Unfortunately, this early attempt at building empirical models with satisfactory steady state operating characteristics had less impact than it might, owing to other features - an emphasis on continuous time and the role of "disequilibrium" money - being accorded more prominence. A new class of models is now emerging - for example Murphy et al. (1986) - which emphasise consistency in actions and which extensively utilise the implicit steady state relations imbedded in each of the dynamic relations to assess the long-run consequences of policy changes and to provide measures for desired or equilibrium quantities like the exchange rate and inflation. As these models are still in their infancy it is too early to pass a definite judgement regarding their lasting impact on econometrics. Nevertheless, I expect it to be substantial.

3.4 Achieving Steady State Consistency

Mentioned above was the growing conviction that models should exhibit sensible steady state solutions. What implication does this proposition have for the design of individual equations? A characteristic of time series modelling is the need to allow for the fact that economic actions are characterised by slow adjustment. Thus, if income increases, it takes time for this to be translated into decisions on consumption. If adjustment was immediate, the observed ratio of consumption to income would be constant, and the variance of income would be expected to be close to the variance of consumption. In fact, for Australian data, whilst the ratio is "constant", the variance of income is some ten times that of consumption. Accounting for this pheonomenon was among the earliest challenges facing econometricians. By

1966 a popular strategy for dealing with it had emerged; the use of the partial adjustment model (PAM):

$$\Delta y_t = (\lambda - 1)y_{t-1} + \beta(1 - \lambda)x_t. \tag{11}$$

If x_t was i.i.d. (μ,σ^2) it is clear that $E(y_t) = \mu\beta$ and so the ratio of $E(y_t)$ to $E(x_t)$ would be β, whilst the ratio of the respective variances of y_t and x_t would be $\beta^2(1 - \lambda)/(1 + \lambda)$. Hence, as λ rose towards unity, the variance of y_t could become very small relative to x_t; exactly what is observed in the data.

Although the PAM captures many of the salient characteristics of the data, it has the disadvantage of forcing adjustment to be greatest in the first period, before declining in a geometric progression. Many proposals were made in the years immediately prior to 1966 to relax the restrictive shape of lag distribution associated with the PAM; Griliches (1967) provided a survey of these suggestions which still remains one of the best treatments of the topic. What developments there were after that time largely constituted refinements of the basic ideas. Shiller (1973) and Poirier (1975) modified Almon's (1965) polynomial distributed lag model by allowing deviations from a fixed polynomial order, while Godfrey and Poskitt (1975) discussed the determination of polynomial order as testing a sequence of nested hypotheses.

Until the mid 1970's little of this research was concerned with overcoming the deficiencies of the PAM by a direct extension of its philosophy. That was unfortunate, particularly in the new climate of concern over steady state features, since one of the most desirable characteristics of the PAM was the fact that, if x_t was set to $\bar{x}, y_t$ would converge to $\beta\bar{x}$. Perhaps it is not surprising then that, in 1978, Davidson et al. re-introduced to econometrics a simple generalisation of the PAM - error correction mechanisms (ECM's). If $y^*_t = \beta x_t$ is the equilibrium or steady state value of y_t, a simple type of ECM would be:

$$\Delta y_t = \alpha_o \Delta y^*_t + \alpha_1 (y^*_{t-1} - y_{t-1}). \tag{12}$$

where $y_t = \log Y_t$, $Y^*_t = \log Y^*_t$

Now the PAM (11) could be re-written as:

$$\Delta y_t = (1 - \lambda)\Delta y^*_t + (1 - \lambda)(y^*_{t-1} - y_{t-1}). \tag{13}$$

showing that the ECM generalises the PAM by not requiring the impact of disequilibrium effects - α_1 - to be equal to those for equilibrium effects - α_o. The "error" is $(y^*_{t-1} - y_{t-1})$ and the "correction" made is $\alpha_1(y^*_{t-1} - y_{t-1})$.

Inspection of (12) reveals that y_t converges to a constant y^*_t as $t \to \infty$ when $| \alpha_1 | < 1$, and hints at the possibility of jointly allowing for a wide variety of adjustment schemes while preserving steady state properties by replicating the format as in (14):

$$\Delta y_t = \sum_{j=1}^{p} \alpha_j \Delta y^*_{t-j} + \sum_{j=1}^{q} \beta_j (y^*_{t-j} - y_{t-j}). \tag{14}$$

Notice that two principles interact in the specification (14): the model is *designed* to reproduce desired steady state behaviour and the data is called upon to provide information on the appropriate adjustment path (p, q, α_j, β_j). Segmentation of these tasks preserves the role of the theorist in specifying equilibrium behaviour and the econometrician in extracting information on dynamics.[8]

Earlier it was mentioned that the "new" macro-econometric models featured a steady state growth *path*. How does this departure from a target *level* to a growth *rate* impinge upon model design. Return to the PAM in (13) assuming that y_t and x_t are the logs of variables Y_t and X_t respectively, so that Δy^*_t is the growth rate in Y^*_t. Then:

$$\Delta(y_t - y^*_t) = -\lambda\Delta y^*_t + (1 - \lambda)(y^*_{t-1} - y_{t-1}), \tag{15}$$

and $y_t - y^*_t$ will not converge to zero unless $\Delta y^*_t = 0$ i.e. when Y^*_t exhibits steady state growth the change in the ratio of Y_t to Y^*_t depends upon this growth rate. Currie (1981) pointed this out, arguing that in many instances it was unacceptable from theory for such "rate of growth effects" to be in evidence in any steady state.

---

[8] Nickell (1985) has shown the ECM format to be an optimal response for certain types of behaviour in x_t. It should also be emphasized that steady state behaviour for stock/flow variables is naturally enforced by the presence of a stock disequilibrium term. Carland and Pagan (1979) impose steady state behaviour upon an output equation by including a term representing the departure of inventories from their equilibrium levels, while Hendry and von Ungern-Sternberg (1979) argue for an "integral effect" of liquid asset balances upon consumption behaviour.

What can be done about this phenomenon? The most popular solution has been to extend the ECM by adding the term $(1-\alpha_o)\Delta y_{t-1}$ to the RHS of (12) to obtain:

$$\Delta y_t = \alpha_o \Delta y^*_t + \alpha_1(y^*_{t-1} - y_{t-1}) + (1 - \alpha_o)\Delta y_{t-1}, \tag{16}$$

$$= -(1 - \alpha_o)\Delta^2 y^*_t + \alpha_1(y^*_{t-1} - y_{t-1}) + (1 - \alpha_o)(\Delta y_{t-1} - \Delta y^*_{t-1}) + \Delta y^*_t \tag{17}$$

where $\Delta^2 y^*_t = \Delta y^*_t - \Delta y^*_{t-1}$.

After subtracting Δy^*_t from both sides of (17) and re-arranging, we get:

$$y_t - y^*_t = (2-\alpha_1-\alpha_o)(y_{t-1} - y^*_{t-1}) - (1-\alpha_o)\Delta^2 y^*_t - (1-\alpha_o)(y_{t-2} - y^*_{t-2}). \tag{18}$$

In steady state growth $\Delta^2 y^*_t = 0$ and $y_t - y^*_t \to 0$ provided α_1 and α_o are such that the difference equation is stable. This is Salmon's (1982) solution because (16) could have been expressed as:[9]

$$\Delta^2 y_t = \alpha_o(\Delta y^*_t - \Delta y_{t-1}) + \alpha_1(y^*_{t-1} - y_{t-1}). \tag{19}$$

Salmon therefore enforces steady-state growth by extending the ECM

---

[9] There is more than one way of getting this result - Pagan (1985) - including adjusting the constant terms in equations, a method favoured in the type of models formulated in Bergstrom (1976). Others reduce to the condition prescribed by Currie (1981) that the "mean lag" response of y_t to y^*_t should be zero.

dynamics from first order in (12) to second order in (18), and some modellers have opted to do this automatically - Murphy et al. (1986). I do not feel this is a good idea, and the implied constraint should at least be tested. When $\Delta y^*_t = \Delta x_t$ i.e. $\beta = 1$ it is only necessary to test if the coefficient of the introduced variable Δy_{t-1} and that of Δx_t sum to unity in (16). Unfortunately, if $\Delta y^*_t = \beta \Delta x_t$ this solution is no longer available and a non-linear restriction must be examined, either as in Patterson and Ryding (1984) or Pagan (1985).

What emerges from the above discussion is that the time series behaviour of x_t should influence the specification of econometric models. Certainly, something needs to be done to guarantee a steady state equilibrium in the face of differing growth paths, and the fact that most econometric models (both of the single equation and systems of equation variety) have ignored the issue in the past is unsatisfactory. Conceivably, part of the poor performance of models once growth rates were disturbed after the oil price shocks of the 1970's might be accounted for in this way.

This interdependence between model specification and the time series behaviour of x_t might be viewed as unfortunate; it certainly makes life much more difficult than it appeared in the mid 1960's. But, oddly enough, non-stationarity in x_t may even be a benefit. To appreciate this, it is useful to adopt a different re-parameterisation of the PAM than the ECM variant, namely, that provided by Bewley (1979):

$$y_t = \beta x_t - (1 - \lambda)^{-1} \lambda \Delta y_t . \tag{20}$$

Although (11) and (20) are identical, the re-parameterisation in (20) proves to be very useful as the coefficient of x_t is the long-run response, whilst that of Δy_t is the mean lag.[10] Suppose now that x_t follows a first order deterministic time trend and that y_t is regressed against x_t, ignoring entirely Δy_t. From regression theory:

$$\hat{\beta} = \beta - (\lambda/(1-\lambda))\ (\Sigma x_t^2)^{-1}\ \Sigma x_t\ \Delta y_t + o_p\ (1), \tag{21}$$

so that $\hat{\beta}$ will be consistent if the second term is $o_p(1)$. But the denominator is $O(T^3)$, the numerator is only $O(T^2)$, making $\hat{\beta}$ consistent i.e. long-run information on the system is available by *totally ignoring* the dynamics. This is separation of tasks with a vengeance! Granger and Engle (1987) first drew attention to this remarkable result, although one can see precursors to it in the literature. Kramer (1982) for example showed that, when β was identified, $\hat{\beta}_{OLS}$ would be consistent even if x_t was an endogenous variable, provided the exogenous variables of the system exhibited time trends, and this is just an application of the idea above in that the cross product between x_t and a stationary error term will be of lower order than Σx_t^2.

Not all the news about non-stationarity is good, however, particularly when x_t follows a non-ergodic linear process such as an ARIMA process. When $x_t = y_{t-1}$ and $\beta = 1$, Dickey and Fuller (1979) observed that the limiting

[10] As mentioned in Hendry and Pagan (1980) there are many other re-parameterisations which could be performed.

distribution of $T(\hat{\beta} - 1)$ was not normal, and used simulation methods to tabulate the empirical density function for $T(\hat{\beta} - 1)$. Since many economic times series are well modelled by ARIMA processes, Nelson and Plosser (1982), this suggests that the limiting distribution of the OLS estimator of β could also be non-normal. In a series of papers, Phillips and Durlauf (1986), Park and Phillips (1986), Stock (1985) and Sims, Stock, and Watson (1986) have gone a long way toward establishing the limiting behaviour of the suitably standardized OLS estimator when x_t is allowed to be stationary or non-stationary. Suppose x_t was ARIMA (p,1,q). Then, distributed lag models of the ECM type in (12) could be re-parameterized as

$$\Delta y_t = \alpha_0 \beta \Delta x_t + \alpha_1 (x_{t-1} - y_{t-1}) + \alpha_1 (\beta - 1) x_{t-1} \tag{22}$$

$$= \alpha_0' \Delta x_t + \alpha_1 (x_{t-1} - y_{t-1}) + \alpha_2' x_{t-1} \tag{23}$$

and the first regressors will be stationary linear processes whilst the last will be non-stationary. Traditional asymptotic theory then applies to the estimators of α_0' and α_1, whilst the distribution of the estimator of α_2' is more complex. It is very hard to provide a short summary of the results from those papers, since much depends on whether the x_t processes have a deterministic trend or not, and whether x_t is a scalar or a vector, and whether the processes trend together in a linear fashion (or are co-integrated in Engle and Granger's (1987) terminology). Notice, however, that for many models β will be unity in (20), as this occurs whenever X_t and Y_t are in constant ratio in equilibrium, and therefore all variables in (23) are stationary. Consequently, the fact that x_t follows an ARIMA process is of no importance for the estimation of the re-parameterized model. Considerations such as this suggest that the ECM format is a very useful way to express dynamic relations in econometrics, and the parameterization that it adopts can also aid the application of standard statistical theory.

4. The Search for Robustness

The conclusions drawn from econometric models depend upon the assumptions made in constructing them. As discussed earlier, a lot of attention was paid to methods for evaluating the adequacy of these assumptions. Instead, it might be desirable to try to develop procedures making any conclusions drawn from the research robust to departures from these assumptions. Johnston, in 1964, commented upon this issue as follows:

> "Unfortunately, the present stage of development of the subject might be likened to a primitive stage in medicine where a doctor is able to treat only one complaint at a time: he can reset a broken arm or prescribe for influenza, but if you come to him with both these troubles at once, the poor fellow is baffled and is forced to select one of your ailments, treat that, and leave the other alone" (p.147).

An account of attempts to produce a "robust" econometrics may be structured as in section 2 by concentrating upon the specification of conditional moments. As will become apparent, the primitive stage of medicine has been left, but no definitive cure has yet emerged for multiple ailments, although the search for one is currently being pursued on a number of research fronts.

4.1 Making Inferences Robust to the Conditional Mean

Specification of the conditional mean should be the function of economic theory. In practice, however, that theory tends to supply a menu of variables to appear in a relation without indicating how the combination should be performed. Early work in econometrics largely ignored this problem, favouring simple models such as linear in logs or levels, largely because the restricted computational facilities made the estimation of more complex formulations very difficult. Perhaps the earliest reactions against this "ad hocery" arose from the interest of econometricians in estimating price and income elasticities for many commodities. As consumer theory is one of the oldest and best developed branches of economics, it was well understood that restrictions must exist between these elasticities to ensure consistency with maximising behaviour.

Translating these theoretical insights into practice was a rich source of theoretical work in the late 1960's. Prima facie, estimation looks simple, involving only a system of seemingly unrelated equations with cross equation restrictions, and that problem had been intensively studied by Zellner (1962) and was familiar from Goldberger (1964). But there was a complication. The dependent variables of the equations summed to a predetermined aggregate, implying that the disturbances had a singular distribution. McGuire et al. (1968) and Powell (1969) showed, however, that this constraint implied that one equation be dropped in estimation, and that the estimation results were invariant to the particular choice of equation. After that, complete systems of demand equations, imposing and testing the restrictions of consumer theory, appeared - Byron (1970), Barten (1969) and Deaton (1974). Byron's paper tested the restrictions with the Lagrangian Multiplier test statistic, introducing the work of Aitchison and Silvey (1958) into econometrics.

Although this literature was important for re-emphasising the role of economic theory in econometrics, it also required arbitrary decisions upon the nature of the underlying production and utility functions. Byron (1970), for example, estimated a system of log linear demand functions, the implicit utility function being linear in the logs of quantities. It is not all surprising then that interest should have arisen in the possibility of beginning with a functional description of the underlying technology that was sufficiently general to yield a good approximation to the actual one. Diewert (1973) introduced the concept of a "flexible functional form" (defined as functions capable of exactly reproducing all derivatives of the unknown function up to second order), and that line of investigation was a persistent theme in production and consumption economics in the 1970's and 1980's. Since many quantities of interest e.g. elasticities of substitution, are functions of the first two derivatives, the basic property of a flexible form was an attractive one.

Deaton (1986) has written an excellent survey of the popular flexible functional forms emerging in demand analysis - the AIDS, Rotterdam and trans-log models. The latter, set out in Christensen et al. (1973) for production technology, makes the log of output y_t a quadratic function of the log of factors x_t:

$$y_t = y_t^* + x_t\alpha + x_t'Bx_t . \tag{24}$$

The translog form (24) was extensively used in empirical research in the last decade - see Jorgenson (1986) for a detailed account. Because α and B are unrestricted (24) constitutes a flexible form, but that says nothing about the extent to which the approximation is good over a region rather than locally, and Wales (1977) found that various curvature properties operating upon an original function were violated by the approximating one. Nevertheless, the trans-log function does come out of many studies as a good method of approximation.

Can it be improved upon? As (24) may be regarded as a Taylor series approximation to a general function, the direction of improvement would seem to reside in expansions other than Taylor's. Barnett (1983) suggested that the Laurent series expansion, which has terms such as x^k and x^{-k} could offer superior approximating properties. Documentation of this point is to be found in Barnett et al. (1985), but it is too early to render a judgement on these claims. An alternative approach is to vary the order of approximation with the sample size. Thus Gallant (1982) and Elbadawi et al. (1983) employ a combination of a quadratic approximation and a Fourier series expansion; the number of terms in the Fourier series being a function of the sample size. As anticipated this strategy eventually recovers the actual conditional mean when sample sizes become very large, provided that data and functions are "well behaved".

What this development portends is a movement towards non-parametric estimation; Gallant's idea of making the degree of parameterisation dependent on sample size being an attempt to get the benefits of not adopting an explicit parameterisation but retaining all the advantages of a parametric form. For many years it has been known that, with a sufficiently large number of observations and i.i.d. data, it was possible to find the expection of y_t conditional upon x_t without specifying the functional form. Rosenblatt (1956) noted that it was possible to consistently estimate a density function at a point by effectively smoothing the empirical density function with a "kernel"; since the regression model specifies a conditional density it is a natural step to co-opt the same theory to estimate the unknown conditional expectation, and this was done by Nadaraya (1964) and Watson (1964). But the method has been litle used in econometrics, probably because of the nature of economic time series and sample sizes.[11] Robinson (1983) showed that the results could be extended to stationary time series obeying certain mixing conditions, and reviews this time series literature in Robinson (1986a). However, it is in the area of data on individual units that the procedure is likely to have its greatest impact, since the complexity and arbitrary nature of the functional forms prescribed go hand in glove with large amounts of data. Moreover, what is central to many of these studies is a query about aspects of the conditional mean.

[11]Stock (1985) estimates the conditional mean in an analysis of environmental policy, whilst Pagan and Ullah (1986) calculate the conditional variance as a measure of risk. Of course, spectral analysis is a case of non-parametric density estimation and that has had quite a bit of use.

4.2 Making Inferences Robust to the Second Moment

Inferences in econometrics generally demand that an estimate of the variance of estimators be made. Traditionally, it was assumed that the error term in regressions, or the score vector in more advanced cases, had a constant conditional variance. Detailed examination of the validity of this assumption has assumed a position of importance in the past twenty years, although it has to be conceded that responses to evidence of any heteroskedasticity are problematic, as it is a very difficult matter to decide upon a reasonable alternative to the constancy assumption. A number of norms have emerged in econometrics - weighting by a regressor, by the conditional mean, or ARCH type processes (Engle (1982)) - but many empirical researchers felt uncomfortable adopting such tightly parameterised solutions.[12] Accordingly, the desirability of making inferences from econometric estimation robust to any type of heteroskedasticity in the innovations is obvious.

Once again, the solution to this problem was known in statistics before its translation into econometrics - Eicker (1967). But it does not seem to have been used in that discipline until White's (1980a) import of it to econometrics. The underlying idea is very simple. From section 1 a wide variety of econometric estimators solve $T^{-1}\Sigma\hat{d}_{\theta,t} = T^{-1}\Sigma d_{\theta,t}(\hat{\theta}) = 0$. A first

---

[12]Recently Robinson (1986b) and Newey (1986) have shown that fully efficient estimators of β in (1) can be constructed when σ_t^2 is unknown by performing weighted least squares with weights $\hat{\sigma}_t^{-1}$, where $\hat{\sigma}_t^2$ is the non-parametric estimate of the conditional variance.

order Taylor series expansion around the true value θ_o yields $T^{1/2}(\hat{\theta} - \theta_o) = A^{-1}T^{-1/2}\Sigma d_{\theta,t} + o_p(1)$, where A is plim $T^{-1}\Sigma\partial d_{\theta,t}/\partial\theta'$ and $d_{\theta,t}$ is evaluated at θ_o. Hence, the variance of $T^{1/2}(\hat{\theta} - \theta_o)$ depends upon $B = E[(\Sigma d_{\theta,t})(\Sigma d_{\theta,t})']$. When the $d_{\theta,t}$ are independent random variables with bounded variance, a consistent estimator of B is $T^{-1}\Sigma\hat{d}_{\theta,t}\hat{d}'_{\theta,t}$. In the least squares case (1), $d_{\theta,t} = x_t'(y_t - x_t\beta)$ and B is estimated by $T^{-1}\Sigma x_t'x_t(y_t - x_t\hat{\beta})^2$, where $y_t - x_t\hat{\beta}$ will be the least squares residuals.

Of course, this result extends out of the class of independent observations. Whenever $d_{\theta,t}$ is a martingale difference with respect to past data, $E(d_{\theta,t}d'_{\theta,t-k}) = 0 \ \forall \ k > 0$, and so $E(T^{-1}(\Sigma d_{\theta,t})(\Sigma d_{\theta,t})') = T^{-1}\Sigma E(d_{\theta,t}d'_{\theta,t})$. With suitable restrictions upon the nature of $d_{\theta,t}$ this expectation can be consistently estimated by $T^{-1}\Sigma\hat{d}_{\theta,t}\hat{d}'_{\theta,t}$. The main impact of such an enlargement of the domain of Eicker's and White's conclusion is to allow x_t to contain lagged values of y_t - see Nicholls and Pagan (1983). There can be little doubt that the proposed adjustment to make regression-type inferences robust to heteroskedasticity was a particularly beneficial one for applied econometrics. Its widespread use since 1980, and the fact that it can be adapted to many different situations, has meant that at least one of the ills of econometrics has found a medicine to at least suppress the symptoms.

4.3 Robustness to Data Dependence

Economic time series data are rarely independently distributed, and inferences will be affected whenever the quantities $d_{\theta,t}$ are dependent.

Sometimes the dependence stems from the theoretical construct, as in the relationship between a future price for three periods into the future $F_{t+3/t}$, and the spot series at that time S_{t+3}. In such cases it is customary to assume that $E(S_{t+3}|\text{data at time } t) = F_{t+3/t}$, making the difference $v_{t+3} = S_{t+3} - F_{t+3/t}$ have the property $E(v_{t+3}|\text{data at time } t) = 0$, which is compatible with v_{t+3} following an MA(2).

Arguments of this sort led to interest in methods for making inferences robust to dependence in $d_{\theta,t}$. What makes the adjustment complex is that $E(T^{-1}\Sigma d_{\theta,t}d'_{\theta,t})$ cannot generally be replaced by $T^{-1}\Sigma E(d_{\theta,t}d'_{\theta,t})$; if $d_{\theta,t}$ was a martingale difference with respect to past data it could, but that is not true for the future/spot rate example above. When $d_{\theta,t}$ exhibits K'th order dependence, $E(T^{-1}\Sigma d_{\theta,t}d'_{\theta,t})$ might be approximated by $T^{-1}\Sigma E(d_{\theta,t}d'_{\theta,t}) + 2T^{-1}\sum_t \sum_{k=1}^{L} \{E(d_{\theta,t}d'_{\theta,t-k}) + E(d_{\theta,t-k}d'_{\theta,t})\}$, after which this expression is quantified by substituting $\hat{\theta}$ for θ.[13] When K is finite, $L \geq K$ gives an exact result, and a variant of this idea appeared in Hansen and Hodrick (1980). However, if K is not finite, certain restrictions must be imposed if the approximation is to consistently estimate the unknown variance. These restrictions state that the dependence in the data must die out in a particular way, and L must tend to infinity but at a slower rate than K. Domowitz and White (1982) pioneered analysis of this issue, imposing "mixing conditions" upon the $d_{\theta,t}$ and setting $L = o(T^{1/3})$. This last restraint turned out to be slightly incorrect. As

[13] Note $\hat{\theta}$ must be consistent and this means that $T^{-1}\Sigma d_{\theta,t} \xrightarrow{P} 0$ is needed, thereby restricting the type of model that this trick is applicable to.

shown by Phillips (1987) and Newey and West (1987), $o(T^{1/4})$ is needed. To date little experience has been accumulated with this procedure, but it represents another promising step on the road to making econometrics more robust to errors in its maintained assumptions.

4.4 Robustness to Distributional Assumptions

Robustness of estimators of β in (1) to the distribution of u_t attracted an increasing amount of attention over the period under survey, and a variety of approaches emerged. Because the OLS estimator is robust to the distribution of u_t asymptotically (predictions made with it are not), one response would be to formulate criterion functions, minimization of which would yield consistent asymptotically normal estimators of β that may have better small sample performance. Bassett and Koenker (1978) represent the most important reflection of this theme in econometrics. In the first paper the limiting distribution of the estimator of β minimizing the absolute deviations $\sum_{t=1}^{T} |y_t - x_t\beta|$ was found. When $x_t = 1$, making β the location parameter for y_t, this estimator is the median. Other popular location estimates are the α-trimmed means $(1-2\alpha)^{-1}\int_{\alpha}^{1-\alpha} F_n^{-1}(u)du$, where F_n is the empirical distribution function of y_t, and the second paper above generalized this idea to regression models, producing "quantile-regression estimators".

An alternative procedure is to find the MLE that would be appropriate if the true density for u_t was known. Denote this by u_t, so that the log likelihood for (1) would be $L = \sum_{t=1}^{T} L_t = \sum_{t=1}^{T} \log f(y_t - x_t\beta)$. Gallant and Nychka (1987) suggested that $f(u_t)$ be approximated by a series of Hermite polynomials, and that the order of the terms in the expansion be increased with T. They prove consistency of the resulting estimator of β.

Suppose we actually knew $f(u_t)$. Then the MLE of β could be obtained by using the method of scoring, which prescribes the iterative scheme

$\hat{\beta}_{(n)} - \hat{\beta}_{(n-1)} = \left[\sum_{t=1}^{T} (\partial L_t/\partial\beta)(\partial L_t/\partial\beta)'\right]^{-1} \sum_{t=1}^{T} \partial L_t/\partial\beta$, where all derivatives are evaluated with $\hat{\beta}_{(n-1)}$, the estimate of β at the (n-1)'th iteration. Since β_{OLS} is consistent, an estimator that is asymptotically as efficient as the MLE could be found by setting $\hat{\beta}_{(0)} = \hat{\beta}_{OLS}$. Now the problem with implementing this solution is that $\partial L_t/\partial\beta = x_t' f^{-1}(u_t)(\partial f(u_t)/\partial u_t)$ and f is unknown. However, since $\hat{u}_t = y_t - x_t\hat{\beta}_{OLS}$ is a consistent estimator of u_t, $f(u_t)$ can be estimated from the empirical density function of $\hat{u}_t$ by some non-parametric method, and the derivative $\partial f(u_t)/\partial u_t$ may also be estimated in the same way (see the survey by Ullah (1987)). This strategy, adaptive estimation, implements a proposal of Stone (1975), the utility of which was investigated extensively for econometrics by Manski (1984).

Instead of forming a non-parametric estimate of $f^{-1}(u_t)\ \partial f(u_t)/\partial u_t$, one might approximate it by a polynomial in u_t. Then the expectation of such terms involves the convolution of x_t with the conditional moments of u_t. Hence, it seems reasonable that a GMM estimator of β exploiting the orthogonality conditions between x_t and the centered moments of u_t could yield an estimator of β that is as efficient as the MLE, provided the number of moments is allowed to expand with the sample size. Newey (1987) has proven that this is so, and given an example of its use. The estimator is clearly an attractive one as it can be implemented on existing GMM software.

By and large though robust estimation did not make such impact upon time-series based econometrics performed with the linear model (1). In some ways this is surprising. By definition, little is known about the error term in (1), and it might be thought desirable to conduct analysis under a minimal set of conventions. Hausman (1982) probably reflects the opinion of many econometricians upon the matter. He notes that many of the examples used to demonstrate the sensitivity of OLS to distributional assumptions represent the error term as a mixture of random variables, and it is not clear why this is a good description of the errors in econometric models. In practice many of the unusual movements in y_t are the product of events such as strikes, wars, etc. which may be better captured with constructed variables than regarded as generated from a contaminated normal distribution. Because of this feature, econometricians have tended to "patch up" the linear model and report results from a well understood alternative. There are exceptions to this rule. For models featuring unit data, where the asymptotic distributions of popular estimators depend upon the actual distribution of the data, and are not independent of it as is true of OLS applied to (1), the design of robust estimators has attracted a lot of attention. Attempts to grapple with that problem will be discussed in the following section.[1]

4.5 Robustness to System Specification

Economic decisions are interdependent, making it hard to avoid the necessity of posing and answering questions within a systems' framework. This requirement can be troublesome. In estimation it would be unfortunate if estimates of parameter values in one part of a system were rendered incorrect owing to errors made in specifying another part. That issue arose with particular force in the modelling of expectations; it may be easy to list some of the information exploited by agents in forming expectations, but very difficult to make the list complete. Acting as if it was complete normally induces inconsistencies into estimators of the effect of expectations - Nelson (1975). To overcome this, estimators are best formulated from orthogonality relations between the known information set and disturbances. In simultaneous equation parlance, that technique was usually referred to as "truncated Two Stage Least Squares", from the fact that a truncated rather than complete set of predetermined variables was employed as instruments. Brundy and Jorgenson (1971) emphasised this partial approach to estimation in a system.

A legitimate query is whether the problems posed by systems are just those of estimation. Skepticism about this point reached its peak in Sims' (1980) paper. Sims argued that economic theory was just not strong enough to

---

[1]Robustness to variation in the data points x_t was also a subject of study, principally by Krasker and Welsch (1982) in their "bounded influence" estimators, which effectively down-weight influential observations. Koenker (1982) argues that this is largely a response to a failure in the linearity assumption. It may well be that these points are the informative ones, e.g. they may come from a new regime, so that automatic re-weighting seems dangerous.

reduce the number of explanators in an equation to the level where estimation was feasible, and hence that the focus of attention should not be upon the structural equations like (7) but "reduced form" equations like (8) and (10). Answering questions by reference to the latter was to be preferred as it safeguarded against false conclusions being reached due to contamination from invalid assumptions made in the specification of the equations constituting the structure.

In fact, Sims proposed for this task, not the classical reduced form, but a high dimensional vector autoregressive process (VAR) for y_{1t} and y_{2t}. Conceptually, (8) and (10) need only be re-defined with x_{1t} and x_{2t} incorporating lagged values of y_{1t}, y_{2t} and any exogenous variables of the system. Sims' arguments and technique are too involved to summarise here. Some feeling for them may be had by observing that (7) imposed restrictions upon the reduced form equations (8) and (10), and so it also implies certain restrictions between VAR coefficients when (8) and (10) are re-interpreted as VAR's. Consequently, it is possible to examine propositions about the structure by formulating them as implications for the VAR's. Opinions vary on the worth of doing so and, as Cooley and LeRoy (1985) point out in their excellent critique of this literature, much depends on the objectives of the investigator. Nevertheless, there is a strong case for at least checking if the restrictions on the VAR implied by (7) actually do hold; essentially an action analogous to the classical testing of over-identifiability. I know of no formal attempt to construct a specification error test in this way, although there are examples of informal comparisons of estimated and derived VAR's.

5. The Search For The Individual

In 1966 it would have been difficult to find a piece of applied econometrics that did not use time series data. But by 1986 some six out of thirty-five chapters of the *Handbook of Econometrics* pertained to issues rising in the estimation of models for individual behavior. This re-orientation was wrought in part, if not totally, by the computer revolution. Confidentiality requirements associated with much unit data collected by official agencies could now be satisfied by releasing data upon magnetic tapes without identifiers, while the large volumes of such data, almost impossible to handle in card format, were easily managed through magnetic storage devices. Moreover, the cost of transmitting, recording, constructing, and analysing large-scale surveys had fallen dramatically as computers became more powerful. This allowed many researchers to collect their own data rather than having to rely upon official data.

Nowhere are these developments more dramatic than in "labor econometrics". From being a rather sleepy minor field in 1966 its vitality in 1986 is evidenced by the two surveys devoted to its output in the *Handbook of Econometrics*. Such rapid development presents difficulties to a surveyer. Unlike the coherence discernible in time-series based econometrics, a product of almost a quarter of a century of textbooks and articles aimed at unification, unit-based econometrics has only recently generated equivalent resources, principally Amemiya (1981), (1985), Maddala (1983), and the *Handbook of Econometrics* entries. All of this leaves a strong impression that the dust has not yet settled enough to enable a proper assessment of what is of lasting value and what ephemereal. Moreover, some of the issues discussed individually below can potentially arise in combination, leading to an almost bewildering diversity of models to be analysed. What follows therefore is even more of a personal selection than my previous sections, and my emphasis on four main themes is unlikely to appeal to all.

5.1. Models for the Individual

Prima facie the analysis of unit level data doesn't seem to demand new techniques. Replacing "t" with "i" (for the i 'th unit)in (1) would allow estimation of the linear model with individual data. There may even be some gains, as the absence of a natural ordering for the data makes issues of dependence in u_i less important. Indeed , in much of this literature it is assumed that the u_i are independently distributed. Of course, the switch from a sample of time-points to one of units may not be entirely happy, as it may not be possible to conceive of the number of units (N) becoming very large e.g. if these are states of a federation, thereby restricting the use of asymptotic theory.

Chronologically, the first difficulties encountered by econometricians came when attempts were made to exploit data upon both time series and individual units (panel data) in a bid to get "large samples" by pooling the two types of data. In that context (1) became

$$y_{it} = x_{it}\beta + u_{it} \qquad i = 1,\ldots, N \; ; \; t = 1,\ldots,T. \quad (25)$$

(25) was soon regarded as inadequate, since it was felt that there were

likely to be "individual-specific" effects. Mundlak (1963) generalized (25) to

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}, \qquad (26)$$

treating α_i as "fixed effects" parameters to be estimated.

Summing (26) over time and forming sample averages $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$ etc., allows the representation

$$\bar{y}_i = \bar{x}_i\beta + \alpha_i + \bar{u}_i \quad , \qquad (27)$$

from which it is evident α_i could be eliminated from (25) by working with mean corrected data i.e. $y_{it} - \bar{y}_i$ and $(x_{it} - \bar{x}_i)$. The fixed effects estimator of β obtained by regressing $y_{it} - \bar{y}_i$ against $x_{it} - \bar{x}_i$ was probably the most common procedure employed for estimation with panel data over the period under review; it was easily implemented via dummy variables, the number of fixed effects could range from unity to N at the choice of the investigator, and it could be augmented to allow for "time-specific" effects. Perhaps the main complication to emerge was that consistency of $\hat{\beta},\hat{\alpha}_i$ required that both N and T tend to infinity.

Basic to the estimation strategy described above was the elimination of the individual effects by centering upon the unit temporal means. Defining F as the matrix with $1 - T^{-1}$ as the diagonal and $-T^{-1}$ on the off-diagonals means that β was estimated by applying OLS to the transformed model $Fy = FX\beta + Fu$, leading to the query of what other choices for F might also be satisfactory. Suppose (26) is first differenced to produce

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it} \quad . \qquad (28)$$

This also eliminates any fixed effects. When T=2 mean correction and the first difference methods coincide since $\bar{y}_i = \frac{1}{2}(y_{i1} + y_{i2})$ and therefore $y_{it} - \bar{y}_i = \frac{1}{2}(y_{it} - y_{it-1}) = \frac{1}{2}\Delta y_{it}$.

There are many transformations that will eliminate fixed effects, e.g. n'th differencing $(n \leq (T-1))$. For each corresponding choice of F, $\hat{\beta} = (X'F'FX)^{-1}X'F'Fy = (Z'X)^{-1}Z'y$, and $\hat{\beta}$ will be consistent provided the variable Z satisfies the conditions to be a valid instrument, in particular that it be asymptotically uncorrelated with the error Fu. There is one obvious instance where it fails. If $x_{it} = y_{it-1}$, the "dynamic fixed effects model", $z_t = x_{i,t-1} - 2x_{i,t} + x_{i,t+1}$ for the first difference procedure, and thus must be correlated with the error. Bhagava and Sargan (1983) provide a comprehensive amount of estimation options in this case.

Rather than treat α_i as fixed a literature grew up - Balestra and Nerlove (1966), Maddala (1971) - which looked at the possibility of making α_i a random variable - $\alpha_i \sim N(o,\sigma_v^2)$ - the "random effects" model. Assuming that α_i and x_{it} are independent, $\alpha_i + u_{it}$ becomes the error term in the regression, and its covariance matrix exhibits a particular pattern. Efficient estimation needs to recognize the non-spherical nature of the errors and this leads to the application of GLS, feasible GLS or M.L.E. Much of this literature was concerned with the relative efficiency of each estimator and with expanding the model to allow for random time effects or serial correlation in the u_{it} - see Amemiya (1985, section 6.6) for a detailed discussion of these developments. In practice, the random effects model has been little used;

perhaps because of the concern, mentioned in Hausman (1978), that the regressors x_{it} cannot be plausibly regarded as independent of the u_{it}.

Although the models described above were advocated for "unit" data, the "units" envisaged were really a collection of agents, rather than individuals in the conventional sense of the word. It is perhaps not surprising that the range of estimators employed in the above developments did not fundamentally depart from those used in time series analysis, although the specific details needed to be worked out, and that proved quite challenging. But once data began to become available upon

"...choice of occupation, marriage partner, or entry into a product market....housing choices into 'sub-standard' or 'standard'....(number of telephone calls)....freight shipment mode." (McFadden, 1985, p1396),

this had to change, as it rapidly became apparent that the observed data on y_i could not be validly treated as a continuous random variable. Rather than *quantitative responses*, these data sets frequently only yielded *qualitative responses*.

Generally the best way to characterize this data is as observations upon an indicator variable. In the simplest schemes choice is dichotomous, and the indicator variable y_i can be arbitrarily assumed to be a binary discrete random variable taking the value unity if a decision is made and zero if it is not. Since the decision is probabilistic, the $\Pr\{Y_i=0\} = F_i$, and it is necessary to describe F_i. This formulation has a long history in biostatistics and psychometrics, and econometricians borrowed substantially from that literature in identifying plausible candidates for F_i. An appealing strategy for doing this is to regard the outcome $y_i = 0$ as occurring whenever some latent variable y_i^* fails to cross a threshold C. (1) may then be employed to model y_i^* as $y^*_i = x_i\beta + u_i$, leading to the determination of F_i = Prob $\{y_i=0\}$ = Prob $\{y_i^* < C\}$ = Prob $\{x_i\beta+u_i<C\}$ = Prob $\{u_i < C - x_i\beta\}$. F_i can then be identified with the cumulative distribution function for u_i. Making F the cumulative normal yields the probit model while the logistic gives the logit, but other choices such as exponential or Student's t might be candidates in the search for robustness. Applied studies have, however, rarely departed from the logit or probit forms.

McFadden (1974) recognised the potential of this framework in econometrics. Defining y_{ij}^* as the utility the i'th individual gets from the j'th alternative action he can take, and treating y_{ij}^* as determined by (1), $y_{ij}^* = x_{ij}\beta + u_{ij}$, a particular action J will be adopted only if y_{iJ}^* exceeds the value of y_{ij}^* for all other j. Clearly the choice between alternative J and any other j depends upon whether $y_{iJ}^* - y_{ij}^* = (x_{iJ} - x_{ij})\ \beta + u_{iJ} - u_{ij}$ exceeds zero, i.e., $\Pr[y_i = 0] = \Pr[y_{iJ}^* - y_{ij}^* < 0] = \Pr[(u_{iJ} - u_{ij}) < -(x_{iJ}-x_{ij})\beta]$. By ascribing densities to the u_{ij} maximization of a random utility function generates the probit or logit models.

As apparent from the details above the framework is applicable to multiple as well as binary choice situations. But it was soon discovered that estimation of the probit model in this context was a forbidding task, whilst that for the logit model was relatively straightforward. Having a utility maximizing perspective then proved to be invaluable, and McFadden (1974) used it to demonstrate that the multinomial logit model would be compatible with utility maximization only if all the errors u_{ij} were independent; the restriction earning the designation "independence of irrelevant alternatives". This is an unhappy assumption, and it spawned a literature concerned with checking its adequacy - Hausman and McFadden (1984) - and with relaxing it. The best known solution allowing for some dependence among the u_{ij} was the nested logit model described in McFadden (1981). In this construct choices between very similar alternatives are made as if governed by a binary logit model, but these similar alternatives are then aggregated into a composite alternative for a binary comparison with any district alternative. A number of applications of this model have appeared - Hausman and McFadden (1984), Small and Brownstone (1982) - but it has proven to be computationally demanding.

The above discussion was rooted in the polar extremes for which data is either continuous or discrete, but a combination of the two types is not unusual. Tobin (1958) was interested in applying (1) to data on household durable expenditure and income, but found that there were both zero and positive expenditures. A zero expenditure clearly represents a qualitative response, the decision not to purchase, whereas a positive value is a commitment to purchase a specific quantity. To represent the data probabilistically therefore requires a combination of the densities for continuous and discrete random variables; the likelihood for $y_1 \ldots y_N$ will be $\prod_{i \in I_o} F_i \prod_{i \in I_1} f_i$, where f_i is the density for u_i in (1), I_o are the points for which only a qualitative response occurs, and I_1 represents the remainder. Although zero represents the censoring point above, the "Tobit" model applies whenever y_i^* must pass some threshold value before an action is taken.

Tobin took the f_i to be normal and Amemiya (1973) provided the seminal treatment of the asymptotic properties of the MLE of β and σ^2 for truncated normal regression models, a class which includes the Tobit model. Since 1970 numerous applications of this sort of model have appeared - see Amemiya (1985, Ch. 10). Most of the variability here came from generalizations that replaced (1) by more than a single equation or which allowed the censoring threshold for one random variable to depend upon another, thereby inducing a type of simultaneity. Amemiya (1985) has proposed that all these contributions be classified into five types depending upon the dimension of y_i^* and the censoring rule employed. Despite this diversity in model form, estimation is rarely by anything but MLE. Thus the interesting aspects of this literature are the applications made rather than the econometric theory employed, although sometimes considerable ingenuity is needed in finding the likelihood.

Perhaps the major concern and challenge to econometric theorists in this area has been one of robustness. Unlike the linear regression context in which the scores derived under a normality assumption continue to have zero expectation udner a range of alternative distributions for u_i, this will rarely be true for discrete random variables, inasmuch as the scores involve $\partial F_i/\partial\beta$, and these are a function of F_i itself. Because of that dependence, mis-specification of F_i leads to a score with non-zero expectation when evaluated with respect to the true probability measure, and so the MLE of β will be inconsistent. A related difficulty stems from the fact that the scores for β and σ^2 do not have zero covariance even when f_i has been selected correctly; mis-specification in either the conditional mean or variance can therefore be expected to affect the consistency properties of the MLE of either set of parameters.

The correct log likelihood to be maximized is

$$\sum_{i=1}^{n} y_i \log(1-F_i) + (1-y_i)\log F_i \tag{29}$$

where F_i is the unknown distribution function of u_i. One might ask therefore if $\log F_i$ can be replaced by a quantity $\bar{F}_i$ (and $\log(1-F_i)$ by $(1-\bar{F}_i)$) such that maximizing (29) with these substitutions would generate a consistent estimator of β. Manski (1975) set $\bar{F}_i = I[x_i\beta<0]$, where I is the indicator function, and termed the resulting optimand the "score function". This terminology is most unfortunate, and would be better described as a "success function" in that it captures the number of successful predictions made if y_i was predicted to be unity whenever $x_i\beta\geq 0$. Restricting the parameter space to $B = \{\beta|\beta'\beta = 1\}$, Manski optimizes by a line search method described in Manski and Thompson (1986) and a complete proof of the strong consistency of the estimator is provided in Manski (1985) provided $F(0) = 0.5$.

An alternative approach would be to consider estimating F_i along with β. For given F_i, the MLE of β is consistent and asymptotically normal; while, for given β, consistent and asymptotically normal density estimators of F_i can be obtained by a variety of non-parametric procedures. This suggests that, by iterating between each single estimator, both β and F_i can be jointly determined. Cosslett (1983) described a suitable estimator of F_i and then demonstrated the consistency of $\hat{\beta}$ obtained by maximizing (29) with F_i replaced by its estimator $\hat{F}_i$.

A parallel but more extensive literature has developed upon the robut estimation of Tobit models, the log likelihood of which is (29) with the first term $y_i \log(1-F_i)$ replaced by $y_i \log f_i$. Fernandez (1986) and Horowitz (1986) used the Kaplan and Meier (1956) estimator of F_i while Duncan (1986) approximates f_i by spline functions. Unfortunately, each of these solutions, as well as those detailed above for qualitative response models, shares a fundamental weakness. The unknown F_i's are being treated as nuisance parameters, and their number tends to infinity with the sample size. Although it is possible to invoke Kiefer and Wolfowitz's (1956) theorem on the consistency of the MLE in the presence of infinitely many incidental parameters to prove consistency of $\hat{\beta}$, no corresponding argument for the limiting distribution of $\hat{\beta}$ is currently available. This failing may make such estimators unattractive to applied researchers whose samples are not large enough to ignore distributional questions.

Given this feature it is not surprising that estimators have been proposed for censored models that aim to optimize something other than the log likelihood in (29). For Tobit models the most influential of these has been Powell's (1986) "symmetrically trimmed least squares" estimator of β which minimizes $N^{-1} \sum_{i=1}^{N} (y_i - \max\{\frac{1}{2} y_i, x_i\beta\})^2$; the "symmetry" deriving from the fact that the observations for which $u_i \leq x_i\beta_0$ are discarded, mirroring the natural censoring which has "discarded" observations $u_i \geq -x_i\beta_0$. Powell demonstrated consistency and asymptotic normality for his estimator. Another promising approach has been Ichimura's (1986) method of finding β to minimize $\sum_{i=1}^{N} (y_i - E(y_i|x_i,\beta))^2$. For a given value of β, $E(y_i|x_i,\beta)$ may be recovered by the non-parametric methods of Nadarya (1964) and Watson (1964), so that a suitable iteration scheme will minimize the sum of squares above. Ichimura's proposal applies to both Tobit and Probit models, and has also been shown to be consistent and asymptotically normal.

Robustness of estimators to the characteristics of the error term is a crucial requirement for models based on unit data, but the derivation of suitable estimators and the isolation of their properties is a complex and demanding task, upon which some of the best econometric theorists have labored in the first half of the current decade. Progress has been made, but the current stock of estimators is distinguished by incomplete characterization of asymptotic properties and the computer-intensity of the associated algorithms. In many ways the situation is reminescent of that in the early stages of the development of simultaneous equations methodology; it was not MLE that eventually became the preferred option of applied workers, but the method of moments estimators 2SLS and 3SLS, and these were not part of the original proposals made by the Cowles Commission members. Perhaps a similar history will one day be evident here. The next decade should be an exciting one for workers in this field, but it would be premature to hazard a prediction about which orientation will eventually prove to be the dominant one.

5.2 Selecting the Individual

Early in elementary statistics courses students are exposed to examples of inferences rendered invalid by a failure to draw a sample randomly, e.g. the telephone poll predicting Roosevelt's defeat. Perhaps it is not surprising then that, once the data to be analysed was determined by an investigator, econometricians were forced to analyse the joint issues of sample selection and estimation. Random selection of a population can create no problems for estimation, and the assumption of independence of u_i in (1) is sometimes justified for individual data on this basis. But there are reasons why one might wish to deviate from this ideal, particularly in the light of cost considerations. If the subject to be studied was the choice by Americans of Belgium as a tourist destination it would be very costly to randomly sample the complete American population, since the sample size would need to be very large to capture a reasonable number of individuals choosing Belgium. A cheaper method is to randomly sample only those who had booked to travel to Europe at travel bureaus. But, this method can have its problems. Cast in terms of (1) individuals in the "travel bureaus sample", may be those with higher than average u_i, and this must therefore be allowed for in estimation

Two generic types of selection are stressed in the literature. In the first sampling is done on the basis of the characteristics x_i in (1), and is naturally referred to as *exogenous* sampling. Here, if the finite choice set is denoted by C and the choice model specifies $P(j|x_i,\beta)$, $j \in C$ being the j'th alternative chosen by the i'th individual, the analyst draws a sample i = 1,..,N from a larger population of size N^* according to some density $g(x_i)$. The likelihood of the observed data becomes $\prod_{i=1}^{N} P(j|x_i,\beta)g(x_i)$. When $g(x_i) = f(x_i)$, the actual density for the characteristics, sampling is random; otherwise, it is stratified. Clearly, the nature of exogenous sampling is important only if $g(x_i)$ is a function of β; if it is not the score with respect to β depends solely upon the $P(j|x_i,\beta)$, and so the nature of $g(x_i)$ is irrelevant.

The tourist case cited above is different in that the selection is based upon the choices of agents - travel to Europe - and not upon the x_i. Hence, Manski and Lerman (1977) called this *choice based* sampling. Here the investigator predetermines the probability of observing the elements in the choice set, $H(j)$, by the way he selects from the population. Within this class, sampling is done at random. Hence the probability of observing an alternative j and characteristics x_i is $P_c(j,x_i) = P_c(x_i|j)\ H(j)$, with the c indicating the conditioning feature that sampling was conducted only among those in the population who choose an alternative j. By Bayes rule $P_c(x_i|j) = P(j|x_i,\beta)p(x_i)|Q(j)$, where $Q(j)$ is the probability that an agent in the population would have adopted alternative j, making the likelihood $\prod_{i=1}^{n} (P(j|x_i\beta)\cdot p(x_i)H(j)|Q(j))$. From this expression it follows that, whilst $p(x_i)$ and $H(j)$ do not normally enter the score, the fact that $Q(j) = \int_{x_i} P(j|x_i,\beta)p(x_i)dx_i$ means the contribution of $Q(i)$ to the likelihood cannot be ignored.

Potentially, there are a number of unknowns in this likelihood, particularly $p(x_i)$ and $Q(j)$. When $Q(j)$ is known the likelihood to be maximized is $\prod_{i=1}^{n} P(j|x_i,\beta)$, but with β obeying the constraint $Q(j) = \int_{x_i} P(j|x_i,\beta)p(x_i)dx_i$. Manski and Lerman (1977) observed that straight MLE was complex and proposed instead that a consistent estimator of β be found by maximizing $\prod_{i=1}^{n} P(j|x_i,\beta)w(i)$ where the weights $w(i) = H(j)Q(j)$. Later work provided rigorous proofs for the MLE when Q(j) was unknown - Manski and McFadden (1981) - and MLE when both $f(x_i)$ and Q(j) are unknown - Manski and McFadden (1981), Cosslett (1981).

From all of this it is apparent that the freedom of an investigator to construct his sample complicates estimation. But even if the sample has not been manipulated by the econometrician, does this mean that it may be regarded as randomly selected? Just as the debate over rational expectations showed that it is not satisfactory to consider agents as reacting passively to any policy changes, rather they will actively optimize in the face of such changes, so too it is necessary to consider that the inherited sample is the outcome of optimizing decisions, and therefore it may be unrepresentative of the population. That is, although the error term in (1) has a zero mean when the complete population is examined, or a random sample of it is chosen, the sample presented to a researcher could well have "high" or "low" u_i individuals deleted from it, as such agents' optimizing decisions remove them from the sample entirely. Such a process is referred to as "self-selectivity". Its consequences are to force the remaining u_i's to have a non-zero mean, so that a regression of observations on y_i against x_i will produce an inconsistent estimator of β.

As might be expected given its origin in individual choice behaviour, self-selectivity is a pervasive phenomenon with economic data collected on units. Initially it was highlighted in wage equations seeking to explain wages earned by women y_i as a function of characteristics x_i from data on wage-earners. This represents a sample from a larger population that includes non-wage earners, i.e. those who do not work. If the participation decision is done optimally, it occurs only if the actual wage exceeds some reservation wage $\bar{y}_i$. Hence, all individuals whose u_i is less than $\bar{y}_i - x_i\beta$ will be "self-selected" out of the sample, and the individuals remaining in the sample will have an excessive number of "high" u_i's. Although Roy (1951) drew attention to this issue in discussing the choice of individuals between two occupations, hunting and fishing, on the basis of comparative advantage, it was the work of Gronau (1974), Lewis (1974) and Heckman(1974) upon the earnings equation described above that brought home the phenomenon with great force to econometricians.

In its greatest degree of generality, when (1) describes a latent variable, self-selectivity is an extreme case of choice based sampling, and could be treated from that perspective. However, more direct analysis exploiting some of its specific features has been the preferred way of proceeding. Accordingly, in the context of the labor econometrics example cited above, if the reservation wage is a function of variables z_i i.e. $\bar{y}_i = z_i\gamma$, then $P(u_i \leq \bar{y}_i - x_i\beta) = P(u_i \leq z_i\gamma - x_i\beta)$. Individuals located in the work force are those for whom u_i exceeds $z_i\gamma - x_i\beta$, and the error term for the sample of wage earners is truncated. Assuming that the underlying error u_i is n.i.d.(o,σ^2), Amemiya (1973) demonstrated that the expected value of the error for those in the sample was $\sigma^2\varphi(\sigma^{-1}(z_i\gamma - x_i\beta))/\Phi(\sigma^{-1}(z_i\gamma - x_i\beta))$, where φ is the standard normal density and Φ is the cumulative standard normal. Knowledge of the mean of this error makes it possible to force the regression on the sample data to have a zero mean by the simple expedient of augmenting the regressors in (1) by $\varphi(\cdot)/\Phi(\cdot)$.

Heckman (1976) used this solution, estimating $\sigma^{-1}\gamma$ and $\sigma^{-1}\beta$ by applying a probit model to data on individuals both in and out of the work force, and then constructing regressors $\hat{\varphi}$ and $\hat{\Phi}$ from these estimates. This two-stage solution is very elegant, is simple to apply, and has found favor in many contexts. Apart from its use of strong distributional assumptions to find the appropriate correction to the mean, its major disadvantage is the need to allow for the fact that $\hat{\varphi}$ and $\hat{\Phi}$ are estimated quantities when calculating the covariance matrix of $\hat{\beta}$, but programs such as LIMDEP now automatically provide this adjustment. The literature on self-selectivity is vast, both in terms of applications and in terms of extending the theory to handle simultaneity, Kenny et al (1979); latent variables, Lee (1979); missing data, Griliches et al (1978); and non-normality, Olsen (1982) and Lee (1982). A comprehensive examination of this literature is beyond this survey. However, Amemiya (1985) and Maddala (1985) provide excellent reviews of this literature.

5.3 The Duration of Events

Once unemployment rose sharply after the first oil price shock, and proved to be unusually persistent, examination of its causes became an important issue among economists. Before long it was realized that the unemployment rate was a stock variable and therefore its behavior was compatible with many different underlying flows. Knowledge of the nature of the flows was crucial to any conclusions drawn from unemployment experience and to the appropriate policy responses. To see why, suppose that there are twelve individuals in the workforce, while an unemployment survey taken each month for a year reveals that one person is unemployed in each of these months, giving a constant 8-1/2% unemployment rate. Now this rate could come about in many different ways. Polar cases are when one person is unemployed for the whole twelve months and when each of the twelve is unemployed for one of the twelve months. In the first case, the average duration of unemployment for an individual rises from one to twelve months over the year, while in the latter it remains at one month. Theories of the cause of unemployment, and responses to it, must therefore have a very clear picture of both the stock and flow dimensions of this series.

In a stationary labor market where the number of persons entering unemployment equals the number leaving, the unemployment rate can be decomposed as the product of the proportion of the labor force entering unemployment during the period, the average duration of unemployment, and the frequency (or number of spells) of unemployment. As a prelude therefore to the analysis of unemployment experience, it is important to estimate the average duration of unemployment. Moreover, many economic theories concerned with entry and exit into unemployment make strong predictions about the determinants of the period of unemployment for an individual. Hence there was a considerable demand for and interest in the analysis of series on duration data. Strengthening this tendency was the fact that statistics existed on a number of other types of duration data besides unemployment, e.g. strikes and employment tenure, and so any techniques developed could be readily transferred.

Statisticians have been concerned with the analysis of duration data in biostatistics and reliability theory for many years, meaning that a well-established pool of techniques could be initially imported into econometrics, as was done by Lancaster (1979) in an early presentation. Basic to this literature was the notion of a hazard function derived as follows. If S is defined as the length of a spell of (say) unemployment and treated as a random variable with cumulative distribution function $F(s) = 1-P(S>s)$, and density $f(s)$, the exit probability out of the unemployed state will be

$$h(s) = f(s)/(1-F(s)). \qquad (30)$$

$h(s)$ is the hazard function and it remains to specify how the exit probability $h(s)$ will vary with individual characteristics and spell length.

In the simplest models $h(s)$ is set to a constant. Estimation is then very simple since (30) can be solved to yield $F(s) = 1-e^{-hs}$ and $f(s) = h(1-e^{-hs})$. With data on completed spell lengths for N individuals $S_1,\ldots,S_N$, the likelihood is just $\prod_{i=1}^{N} f(s_i)$, and the unknown parameter can be estimated by maximizing the likelihood.

But it is unlikely that $h(s)$ can be taken to be constant for data on economic agents. First, it would be expected that the escape probability from unemployment should depend upon individual characteristics such as age, education, sex, etc. A minimal re-specification for the hazard function would be to allow it to vary with individuals - $h_i = h(x_i\beta)$ - where x_i is a vector of characteristics. This modifies the density to a conditional one $f(s_i|x_i) = h(x_i\beta)(1-e^{-h(x_i\beta)s_i})$. Estimation by MLE is still fairly straightforward and, as the references in Amemiya (1985) and Heckman and Singer (1986) attest, the most favored strategy. In fact, since $f(s_i|x_i)$ is a conditional density, and $h(x_i\beta)$ is unknown, it might be more appropriate, when N is large, to estimate the density by non-parametric methods, particularly the kernel estimator mentioned in Section 4.1. Although this density may be poorly estimated by such procedures for regions where few x_i are available, if interest is really in statistics such as average duration (E(S/X)), nonparametric methods are probably the most effective way of estimation.

Where problems really arise, however, is when h is allowed to be a function of s - the possibility that the exit probability depends upon the period of time spent in the state - and a recognition that individuals may have differing hazard rates due to unobserved characteristics (neglected heterogeneity). Both of these complicate estimation immensely and research has been devoted to finding expressions for $h_i(s)$ that are both tractable and flexible.

Many proposals have been made to allow the hazard to vary with s. In his germinal piece on duration analysis, Lancaster (1979) specified h(s) as $\alpha s^{\alpha-1}\exp(x_i\beta)$, while Flinn and Heckman (1982) allow for a general Box-Cox transformation on s, thereby relating h(s) to $(s^{\lambda}-1)/\lambda$. Other suggestions have been polynomials in s, ranging from a quadratic - Tuma (1976) - to ninth order - Kennan (1985). Just as in the production function literature detailed in Section 4.1, there is a continuing search for more flexible ways of modelling the dependence of h(s) upon s, and some cross fertilization from the earlier literature might be profitable e.g. Gallant's Fourier series expansions appeal as a flexible form. Non-parametric methods could also be adapted, but if the dependence is complex large numbers of observations will be required.

This is clearly an area with an evident need to investigate the adequacy of any assumed form of duration dependence, i.e. $\partial h/\partial s$. Since estimation has always been done by MLE, diagnostic tests are simply set up by the score test or Newey/Tauchen approaches mentioned in section 2(v). Kiefer (1985) represents one of the few attempts to construct standards in this area. He takes the maintained hazard function to be $h_i = \exp x_i\beta$, giving $f(s_i) = h_i \exp(-s_i h_i)$, and then considers the class of alternative densities $f^*(s_i) = f(s_i)(1+\sum_{j=2}^{n} \alpha_j L_j((\exp x_i\beta)s_i))$, where $L_j(\cdot)$ are the j'th Laguerre polynomials. His alternative covers the exponential, gamma, Weibull and Pareto alternatives for $f^*(s_i)$. Both the gamma and Weibull distributions have had use - Salant (1977), Lancaster (1985). Kiefer constructs LM tests for $\alpha_2,..,\alpha_n$ being zero, based on the restricted scores

$$\left.\frac{\partial \log \sum_{i=1}^{n} f^*(s_i)}{\partial\alpha_j}\right|_{\alpha_j=0} = L_j(\exp(x_i\beta)s_i) \qquad (j=2,...,n) \qquad (28)$$

A much more difficult effect to control for has been unobserved heterogeneity, and attempts to do so have been a major concern of researchers. Neglect of individual heterogeneity has been proven to induce a negative bias into estimates of duration dependence, the most precise statement of this contention being Heckman and Singer ((1986), p. 1205). Lancaster (1979) recognized the issue and wrote the hazard rate as the product of $\alpha s^{\alpha-1}\exp(x_i\beta)$ and a random variable v_i that was identically and independently distributed as a gamma $(1,\sigma^2)$ random variable. To get the density of s_i he integrated out the v_i, and this strategy has been followed by others, although sometimes replacing the gamma distribution with the beta. Unfortunately, as Heckman and Singer (1986, p1710) emphasize, parameter estimates are very sensitive to the specification of a density for v_i. and the need for non-parametric analysis here is paramount. Heckman and Singer (1984) adapted Kiefer and Wolfowitz's

(1956) non-parametric MLE for mixtures of distributions. What is at issue here is that treating u_i as random with unknown density induces an infinite number of incidental parameters into the model. Once again, this makes it hard to establish limiting distributions and it might be better to employ less efficient estimators e.g. one might follow Ichimura and minimize $\Sigma(s_i-E(s_i/x_i))^2$ with respect to α and β.

If one is to stick with parametric procedures, checks for their adequacy would seem mandatory. Lancaster (1985) has been very active in the design of diagnostic tests for such a task, deriving a score test for $\sigma^2 = 0$ against an approximation to the true density suggested by a Taylor series expansion. His test is a very simple one, being based on the generalized residuals, which in his duration model are identified with the integrated hazard function evaluated at the MLE's under the null hypothesis.

In practice, it is even harder to estimate duration models than described above, since data is rarely available on completed spells. Incomplete spells, with unknown initiation (left-censoring) and termination (right-censoring) dates are the norm rather than the exception. Adjustment of the likelihood for this effect is feasible, but it means that a conversion must be effected from the estimated density of incomplete spells to that for completed ones.

Faced with all of these difficulties, as well as the presence of multiple as well as single spell data, official applications of duration data have tended to use the simplest possible models. Thus, statistics on average completed duration have normally been derived by assuming constant hazard functions and gamma-distributed heterogeneity. Probably this will change as software to estimate the more complex models becomes generally available, but at the moment the very youth and vigor of this area of econometric research has been a powerful inhibitor to the widespread adoption of the techniques that it has thrown up.

5.4 From Micro to Macro

The emergence of data upon individual units not only changes the range of models needed in the tool-kit of an econometrician; it fundamentally changed the way in which enquiry could proceed. Because many issues are inherently micro-economic in nature e.g. tax reform and the impact of employment subsidies, use of aggregate data to investigate them misses the diversity in responses of individuals. It might even be argued that the diversity of responses is paramount, as it is the marginal rather than the average response revealed in aggregate figures that an economist seeks. Accordingly, the impact of tax reform upon labor supply may be very strong upon individuals with high marginal tax rates, but very hard to detect in any aggregate data where such individuals constitute only a small fraction of the population. Moreover, data on individuals can supply data sets rich in different environmental conditions, and studying behaviour for the total range might be very informative for the generation and calibration of theories.

Thus there is a strong case for utilizing micro data to investigate macro issues, and research has flourished with the ain of doing just that. By far the most ambitious effort in this direction was initiated by Orcutt (1957) and extended in Orcutt et al (1961), (1976). In these reports the micro decisions of individuals are determined by the conditional probabilities of action;

information that the models discussed in section 5.1 were designed to provide with the conditional probability known, a random draw is made to determine the actual outcome for any individual. Summation of individual histories is then performed to yield outcomes upon macro aggregates. The framework is an appealing one, although in its present version it tends to ignore general equilibrium considerations, mainly concentrating upon supply side decisions and sociological phenomena. A tax reform, for example, would generate effects upon prices and quantities in many markets, whereas in most of these models the computed effects would largely relate to labor/leisure choices, participation decisions etc. Undoubtedly, these strands could be integrated, but no successful variant has yet appeared.

The micro-simulation approach effectively sidesteps what has always been a conundrum in economics; optimizing theory is about an individual but many of the interesting questions are about aggregates. Most macro-econometric analysis has always invoked the "representative agent", but the availability of data on units means an opportunity to explore more rigorous ways of moving from the individual to the "state". Micro-simulation does this numerically, and thereby avoids a set of questions about the process of aggregation that have always bothered econometricians: Should data be pooled under the assumption of a common underlying model? What meaning can be attached to parameter estimates from a pooled sample? How is a macromodel calibrated from time series data build up from those appropriate at a lower degree of disaggregation? Each of these queries has stimulated a modicum of interesting and useful research in the last two decades.

If (1) was such that β varied with individuals, i.e., $y_i = x_i\beta_i + u_i$, pooling all the data in a single regression would be illegitimate. But estimating (1) with dummy variables to allow for each slope coefficient to shift leaves too many unknown parameters. Zellner (1969) suggested a way out of this impasse, by treating the coefficients β_i as realizations from a p.d.f with mean $\bar{\beta}$ and variance V i.e. $\beta_i = \beta + v_i$ where $v_i \sim (0,V)$. Then $y_i = x_i\beta + (u_i + x_i v_i)$, a linear model with heteroskedasticity, to which OLS can be applied to find a consistent estimator of $\bar{\beta}$, after which various estimators of V - Hildreth and Houck (1968), Swamy (1970) - can be computed from the regression residuals. This formulation, known as the random coefficients regression model, is essentially a generalization of the random effects model; the latter could be interpreted as arising from a randomly varying intercept. Zellner proposed that a test of V=0 would be a suitable test of aggregation bias, and this procedure has had some use.[15]

A promising attack was made on the second question in a series of papers by Stoker (1984), (1986). Suppose that $y_i = F(x_i, \beta)$ is the underlying micro-relationship and that x_i is distributed across agents as $p(x|\tau)$. Macro-economic theory might be regarded as being concerned with the response of $\mu_y = E(y_i)$ to a change in $\mu_x = E(x_i)$, i.e. $\partial\mu_y/\partial\mu_x$, as this captures the idea of the representative or average reaction. Can this information be extracted

[15]The idea of treating the coefficients as characterized by a p.d.f. over individuals was also exploited by Trivedi (1985) in studying how lag distributions such as that of Koyck's can arise as a result of aggregation, even though individual units' dynamic behavior is not governed by such simple forms.

from a regression of y_i against x_i? By the chain rule $(\partial\mu_y/\partial\mu_x) = (\partial\mu_x/\partial\tau)^{-1}(\partial\mu_y/\partial\tau)$, provided $\partial\mu_x/\partial\tau$ is non-singular. Now, $\int x\ \partial p/\partial\tau dx = \int x(\partial \log P/\partial\tau)p(x)dx = E(dx) = V_{dx}$ where d is the score $\partial \log p/\partial\tau$ and V_{dx} is the covariance of d with x. Similarly $\partial\mu_y/\partial\tau = V_{dy}$, making $\partial\mu_y/\partial\mu_x = V_{dx}^{-1}V_{dy}$, which is asymptotically equivalent to the IV estimator of the coefficient of x_i in the linear model relating y_i to x_i and where d_i is the instrument used. When $p(x|\tau)$ is a member of the exponential family the IV estimator just becomes the OLS estimator. Hence, a sensible macro-economic interpretation can be given to a linear cross-section regression even though the underlying micro-relationship is a non-linear one that is not preserved in the aggregate data. To be sure, special assumptions are needed to get this outcome, principally that all variation in θ can be ascribed to μ_x, but Stoker's focus upon what can be learned about the average response is an important contribution to our understanding of what can be learned from aggregate data. It is interesting to note that, in his example in Stoker (1986), the OLS estimator gives a good estimate of $\partial\mu_y/\partial\mu_x$, even though $p(x|\theta)$ is not in the exponential family.

A final attempt to build a bridge between macro and micro analysis comes if it is assumed that functions aggregate and an examination is made of how a macro relation estimated from time series would be build up from a series of micro relations. Panel data is an ideal way of doing this; if the sample size is large enough aggregating over all individuals in the panel should generate time series resembling those in the social accounts. Thereupon, by varying the level of aggregation, it is possible to see how much information is hidden by the time series regressions.

Unfortunately, panel data are rare, and unlikely to be consistently available for long periods of time. By their very nature statistical agencies are happier collecting randomly chosen samples from a population in each time period rather than attempting to trace the same set of individuals over time. What exists then are repeated surveys upon the same actions of a different group of individuals over time. Household expenditure surveys are the classic illustration. In this data individuals cannot be tracked but cohorts can, i.e. we can follow the actions of the "age 21-25" cohort as it ages through its representative groups. Estimation proceeds in much the same way as in the fixed effects model, and Deaton (1985) has a comprehensive discussion of how this is done, dealing particularly with the statistical problems arising from the possession of sample rather than population values for each cohort. More work utilizing this framework would seem desirable if we are to fully understand many macro-economic phenomena.

6. The Achievements: A Personal View

How do we take the measure of the achievements of econometrics over this period? For a discipline aiming to provide tools for the analysis of data, some productivity measure is appropriate. To what extent then have the econometric theorists' work of the past twenty years become part of the process of economic investigation and the training of an economist?

Based on this criterion econometrics would have to be credited as an outstanding success. It is hard to find an official report or inquiry by a private or academic group that does not use many, or most, of the techniques discussed above whenever a model is appropriate. Econometric tools are routinely used in forecasting and policy assessment tasks, and a majority of these were developed in the past twenty years.

One of the problems in being more precise about this feature is the fact that the techniques and methods of econometrics have become part of the milieu. It is easy to forget that procedures readily understood and used by undergraduate students were once dimly understood and hotly debated by some of the best minds of the day.

The judging of achievement inevitably involves contrast and comparison. Over a period of twenty years this would be best done by interviewing a time-travelling economist displaced from 1966 to 1986. I came into econometrics just after the beginning of this period, so have some appreciation for what has occurred. But because I have seen the events gradually unfolding, the effects upon me are not as dramatic. Nevertheless, let me try to be a time-traveller and comment on the perceptions of a 1966'er landing in 1986. My first impression must be of the large number of people who now have enough econometric and computer skills to formulate, estimate and simulate highly complex and non-linear models. Someone who could do the equivalent tasks in 1966 was well on the way to a Chair. My next impression would be of the widespread use and purchase of econometric services in the academic, government, and private sectors. Quantification is now the norm rather than the exception. A third impression, gleaned from a sounding of the job market, would be a persistent tendency towards an excess demand for well-trained econometricians. The economist in me would have to acknowledge that the market judges the products of the discipline as a success.

Yet, amidst all this light, there are dark corners. Ted Hannan, in an interview I did with him for *Econometric Theory* in 1985, said about the Cowles Commission program "In fact, its been a little bit of a failure, I'm afraid." Equivalent statements appear from time to time in the financial press or as presidential addresses by eminent figures. Some of this seems sour grapes, but in other cases it runs deeper. There is a nagging feeling that current econometric practice is in some way deficient and that its approach distracts attention away from the economic history and data issues that should be paramount. "In house" criticisms have centered upon whether the probability model adopted by the Cowles Commission is well designed for the types of analysis we typically are forced to do. Leamer (1983) has been a vigorous promoter of this view, and much of that philosophy was shared by the econometricians at CORE in their research upon Bayesian econometrics and statistics - Dreze and Richard (1984). By and large this remains a minority view, not because the philosophy does not appeal, but because the preferred alternative has been more conceptual than practical. I feel very strongly that the failure of CORE's admirably conceived Bayesian Regression Program stemmed from the fact that it was a very unfriendly piece of software. Leamer's greater success with his SEARCH program is attributable, in my opinion, to its greater flexibility and ease-of-use. It will be interesting to see if Bayesian econometrics has a much greater impact if programs can be designed so as to make analysis in this framework relatively easy. Here the interactive capabilities of the personal computer would seem to be a fertile base upon which to build an alternative style of doing econometrics.

"Outsider" reservations are based on a perspective which characterizes econometric theory as a search for optimal procedures in a hypothetical environment that doesn't capture the actual one or, worse, fails to address the essential economic issues. My empathy with some aspects of this belief has steadily grown in the last decade. Econometrics, which began with people who were excellent mathematicians, developed a startling lack of rigor in the 1950's and 1960's, prompting Malinoaud's letter on the topic to the editor of the *International Economic Review* in 1971. Since then we have swung the other way, perhaps too far. Today there is a growing group of econometricians who would appear to regard their peer group as mathematical statisticians, and whose reading is more likely to be the *Annals of Statistics* than the *American Economic Review*. In the hands of extremely skilled people, such as Peter Phillips and Hal White, a lot can be, and has been, learned from this orientation, but it is nowhere near as effective or useful when done by those with lesser skills. Moreover it is worrying that the group of econometricians who can double as reputable economists - for example Jerry Hausman, Jim Heckman and Angus Deaton - does not seem to be gaining new recruits. Associated with this tendency has been the virtual demise, to the detriment of the profession, of the old-style economic statistician who was concerned with index numbers, national accounts, etc. Because of the disappearance of such individuals these topics, prominent in courses twenty years ago, have faded out of the curriculum. "Theory wise, data foolish" is sometimes all too true of our graduates today.

In some ways the notion of achievement is a very personal one, and the best indicator is how the participants feel about it twenty years later. These twenty years for me were exciting times, and I am glad to have made some contribution to the range of tools currently available to assist economic modelling. I firmly believe that econometrics weathered the storms created by flexible exchange rates, supply side shocks, money market de-regulation etc. much better than I would have expected if told of the likelihood of these events in 1966, and it has produced creative responses to them. This paper has tried to exposit those responses. Much remains to be done, and it is that which should make the next twenty years just as exciting and productive as the last.

REFERENCES

Aitcheson, I. and S.D. Silvey (1958), "Maximum Likelihood Estimation of Parameters Subject to Restraints", *Annals of Mathematical Statistics*, 29, 813-828.

Almon,S. (1965), "The Distributed Lag Between Capital Appropriations and Expenditures", *Econometrica*, 33, 178-196.

Amemiya, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal", *Econometrica*, 41, 997-1016.

Amemiya, T. (1981),"Qualitative Response Models: A Survey", *Journal of Economic Literature*, 19, 1483-1536

Amemiya, T. (1985), *Advamced Econometrics* (Harvard University Press, Cambridge, Massachusetts).

Andrews, D. (1985), "Random Cell Chi-Square Diagnostic Tests for Econometric Models: I. Introduction and Applications", *Cowles Foundation Discussion Paper No. 762*.

Avery,R.B., L.P.Hansen and V.J.Hotz (1983), "Multiperiod Probit Models and Orthogonality Condition Estimation", *International Economic Review*, 24, 21-35.

Balestra, P. and M. Nerlove (1966), "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas", *Econometrica*, 34,585-612.

Barnett, W.A. (1983), "New Indices of Money Supply and the Flexible Laurent Demand System", *Journal of Business and Economic Statistics*, 1, 7-23.

Barnett, W.A., Y.W. Lee and M.D. Wolfe (1985), "The Three Dimensional Global Properties of the Miniflex Laurent, Generealized Leontief, and Translog Flexible Functional Forms", *Journal of Econometrics*, 30, 3-31.

Barten, A.P. (1969), "Maximum Likelihood Estimation of a Complete System of Demand Equations", *European Economic Review*, 1, 7-73.

Bassett, G.W. and R.W. Koenker (1978), "The Asymptotic Distribution of the Least Absolute Error Estimator", *Journal of the American Statistical Association*, 73, 618-622.

Bera, A.K. and C.M. Jarque, (1981), "An Efficient Large Sample Test for Normality of Observations and Regression Residuals", Working Papers in Economic and Econometrics No. 049, Australian National University.

Bera, A.K., C.M. Jarque, and L.F. Lee (1984), "Testing for the Normality Assumption in Limited Dependent Variable Models", *International Economic Review*, 25, 563-578.

Berenblut, I.I. and G.I. Webb (1973), "A New Test for Autocorrelated Errors in the Linear Regression Model", *Journal of the Royal Statistical Society* B, 35, 33-50.

Bergstrom, A.R. (ed.) (1976), *Statistical Inference in Continuous Time Economic Models*, Amsterdam: North-Holland.

Bewley, R. (1979), "The Direct Estimation of the Equilibrium Response in a Linear Dynamic Model", *Economic Letters*, 3, 357-362.

Bhagava, A. and J.D. Sargan (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Periods", *Econometrica*, 51, 1635-1659

Bowman, K.O. and L.R.Shenton (1975), "Omnibus Contours for Departure from Normality based on $\sqrt{b_1}$ and b_2", *Biometrika*, 62, 341-348.

Brainard, W.C. and J. Tobin (1968), "Pitfalls in Financial Model Building", *American Economic Review*, 58,

Braynton, F. and E. Mauskopf (1985), *The MPS Model of the United States Economy* (mimeo, Board of Governors of the Federal Reserve System).

Breusch, T.S. and L.J. Godfrey (1986), "Data Transformation Tests", *Economic Journal*, 96, Supplement, 47-58.

Breusch, T.S. and A.R. Pagan (1979), "A Simple Test for Heterscoedasticity and Random Coefficient Variation", *Econometrica*, 47, 1287-1294.

Byron, R.P. (1970), "The Restricted Aitken Estimation of Sets of Demand Relations", *Econometrica*, 38, 816-830.

Byron, R.P. (1974), "Testing Structural Specification Using the Unrestricted Reduced Form", *Econometrica*, 42, 869-884.

Brundy, J.M. and D.W. Jorgenson (1971), "Efficient Estimation of Simultaneous Equations by Instrumental Variables", *Review of Economics and Statistics*, 53, 207-224.

Carland, D.J. and A.R. Pagan (1979), "A Short-Run Econometric Model of the Japanese Wool Textile Industry", *Economic Record*, 55, 317-327.

Challen, D.W. and V.W. Fitzgerald (1984), "Dynamical Features of the NIF-10S Model", in *Proceedings of the Conference on the* NIF-10 *Model* (Australian Government Publishing Service), 186-233.

Chesher, A.D. and M. Irish (1984), "Residuals and Diagnostics for Probit, Tobit and Related Models", Discussion Paper No. 84/152, University of Bristol.

Chow, G.C. (1983), *Econometrics* (McGraw-Hill, New York).

Christensen, L.R., D.W. Jorgenson and L.J. Lau (1973), "Transcendental Logarithmic Production Functions", *Review of Economics and Statistics*, 55, 28-45.

Cooley, T.F. and S.F. Le-Roy (1985), "Atheoretical Macroeconomics: A Critique", *Journal of Monetary Economics*, 16,283-308.

Cosslett, S.R. (1981), "Maximum Likelihood Estimation for Choice Based Samples", *Econometrica*, 49, 1289-1316.

Cosslett, S.R. (1983) "Distribution-Free Maximum Likelihood Estimator of the Binary Choice Model", *Econometrica*, 51, 765-782.

Cox, D.R. (1961), "Tests of Separate Families of Hypotheses", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 105-123.

Cox, D.R. (1962),"Further Results on Tests of Separate Families of Hypotheses", *Journal of the Royal Statistical Society* B, 24, 406-424.

Cox, D.R. and E.J. Snell (1968), "A General Definition of Residuals", *Journal of the Royal Statistical Society*, B, 30, 248-275.

Cumby, R.E., J. Huizinga and M. Obstfeld (1983), "Two-Step Two-Stage Least Squares Estimation in Models with Rational Expectations", *Journals of Econometrics*, 21, 333-335.

Currie, D. (1981), "Some Long-Run Features of Dynamic Time Series Models", *Economic Journal*, 91, 704-15.

Davidson, J.E.H., D.F. Hendry, F. Srba and J.S. Yeo (1978), "Econometric Modelling of the Aggregate Time-Series Relationships Between Consumers Expenditure and Income in the United Kingdom", *Economic Journal*, 88, 661-692.

Davidson R. and J.G. MacKinnon (1981), "Several Tests for Model Specification in the Presence of Alternative Hypotheses", *Econometrica*, 49, 781-793.

Davidson, R. and J.G. MacKinnon (1984), "Convenient Specification Tests for the Logit and Probit Models", *Journal of Econometrics*, 25, 241-262.

Deaton, A.S. (1974), "The Analysis of Consumer Demand in the United Kingdom, 1900-1970", *Econometrica*, 42, 341-367.

Deaton, A. (1985), "Panel Data from Time Series of Cross-Sections", *Journal of Econometrics*, 30, 109-126.

Deaton, A.S. (1986), "Demand Analysis", in Z.Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol 3 (North-Holland, Amsterdam), 1768-1839.

Dempster, A.P., N.M. Laird and D. B. Rubin (1977), "Maximum Likelihood from Incomplete Data Via the E-M Algorithm", *Journal of the Royal Statistical Society*, Series B, 39, 1-22.

Dixon, P.B., B.R. Parmenter, J.M. Sutton and D.P. Vincent (1982), ORANI: A *Multisectoral Model of the Australian Economy* (North-Holland, Amsterdam).

Dhrymes, P.J. et al (1972), "Criteria for Evaluation of Econometric Models", *Annals of Economic and Social Measurement*, 1, 291-324.

Dickey, D. A. and W. A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Diewert, W.E. (1973), "Functional Forms for Profit and Transformation Functions", *Journal of Economic Theory*, 6, 284-316.

Domowitz, I. and H. White (1982), "Misspecified Models with Dependent Observations", *Journal of Econometrics*, 20, 35-58.

Dreze, J.H. and J.F. Richard (1984), "Bayesian Analysis of Sinultaneous Equation Systems", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol I (North Holland, Amsterdam), 517-598.

Duncan, G.M. (1986), "A Semi-Parametric Censored Regression Estimator", *Journal of Econometrics*, 31, 5-34.

Durbin, J. (1954), "Errors in Variables", *Review of the International Statistical Institute*, 22, 23-32.

Durbin, J. (1970), "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables", *Econometrica* 38, 410-421.

Durbin, J. and G. Watson (1971), "Testing for Serial Correlation in Least Squares Regression, III", *Biometrika*, 58, 1-19.

Eicker, F. (1967), "Limit Theorems for Regressions with Unequal and Dependent Errors", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, 59-82.

Elbadawi, I.A., A.R. Gallant and G. Souza (1983), "An Elasticity can be Estimated Consistently Without a Priori Knowledge of Functional Form", *Econometrica*, 52, 1731-1751.

Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflations", *Econometrica*, 50, 987-1007.

Engle, R.F., D.F. Hendry and J.F. Richard (1983), "Exogeneity", *Econometrica*, 51, 277-304.

Engle, R.F., D.M. Lillien and R.P. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, 55, 391-408

Engle, R.F. and C.W.J. Granger (1987), "Dynamic Model Specification with Equilibrium Constraints: Co-Integration and Error Correction", *Econometrica*, 55, 251-278.

Evans, M. and M.L. King (1985), "A Point Optimal Test for Heteroscedastic Disturbances", *Journal of Econometrics*, 27, 163-178.

Fernandez, L. (1986), "Non-Parametric Maximum Likelihood Estimation of Censored Regression Models", *Journal of Econometrics*, 32, 35-57.

Flinn, C. and J. Heckman (1982), "Models for the Analysis of Labor Force Dynamics", in R. Basmann and G. Rhodes (eds.) *Advances in Econometrics*, (Academic Press), 1, 35-95.

Gallant, A.R. (1982), "Unbiased Determination of Production Technologies", *Journal of Econometrics*, 20, 211-245.

Gallant, A.R. and D.W. Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation", *Econometrica*, 55, 363-390.

Godfrey, L.G. (1978a), "A Note on the Use of Durbin's h Test when the Equation is Estimated by Instrumental Variables", *Econometrica*, 46, 225-228.

Godfrey, L.G. (1978b), "Testing for Multiplicative Heteroskedasticity", *Journal of Econometrics*, 8, 227-236.

Godfrey, L.G. and M.H. Pesaran (1983), "Small Sample Adjustments for the J-Test", Working Papers in Economics and Econometrics, No. 084, Australian National University.

Godfrey, L.G. and D.S. Poskitt (1975), "Testing the Restrictions of the Almon Technique", *Journal of the American Statistical Association*, 70, 105-108.

Goldberger, A.S. (1964), *Econometric Theory* (John Wiley, New York).

Goureiroux, C., A. Monfort and A. Trognon (1983), "Testing Nested or non-Nested Alternatives", *Journal of Econometrics*, 21, 83-115.

Gourieroux, C., A. Monfort and A. Trognon (1984), "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 42, 681-700.

Gourieroux, C., A. Monfort and A. Trognon (1985), "A General Approach to Serial Correlation", *Econometric Theory*, 1, 315-340.

Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37, 424-438.

Granger, C.W.J. and R.F. Engle (1985), "Dynamic Model Specification with Equilibrium Constraints: Co-Integration and Error Correction" (mimeo, University of California at San Diego).

Griliches, Z. (1967), "Distributed Lags: A Survey", *Econometrica*, 35, 16-49.

Griliches, Z., B.H. Hall and J.A. Hausman (1978), "Missing Data and Self-Selection in Large Panels", *Annals de L'Insee*, 30-31, supplement, *The Econometrics of Panel Data*, 137-176.

Gronau, R. (1974), "Wage Comparisons - A Selectivity Bias", *Journal of Political Economy*, 82, 1119-1143.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50, 1029-1054.

Hansen, L.P. and R.J. Hodrick (1980), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis", *Journal of Political Economy*, 88, 829-853.

Hansen, L.P. and K.J. Singleton (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models", *Econometrica*, 50, 1269-1286.

Harrison, G.W. (1985), "Sensitivity Analysis and Trade Policy Modelling", in T.N. Srinivasan and J. Whalley (eds.) *General Equilibrium Trade Policy Modelling* (M.I.T. Press, Cambridge).

Hausman, J.A. (1978), "Specification Tests in Econometrics", *Econometrica*, 46, 1251-1272.

Hausman, J.A. (1982), "Comment", *Econometric Reviews*, 1, 271-277.

Hausman, J.A. and D. McFadden (1984), "Specification Tests for the Multinomial Logit Model", *Econometrica*, 52, 1219-1240.

Heckman, J. (1974), "Shadow Prices, Market Wages and Labour Supply", *Econometrica*, 42, 679-694.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement*, 5, 475-492.

Heckman, J. (1984), "The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata", *Econometrica*, 52, 1543-1547.

Heckman, J. and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, 52, 271-320.

Heckman, J. and B. Singer (1986), "Econometric Analysis of Longitudinal Data", in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol 3 (North-Holland, Amsterdam), 1689-1763.

Hendry, D.F. (1980), "Econometrics - Alchemy or Science", *Economica*, 47, 387-406.

Hendry, D.F. and J.F. Richard (1982), "On the Formulation of Dynamic Models in Econometrics", *Journal of Econometrics*, 20, 3-33.

Hendry, D.F. and A.R. Pagan (1980), "A Survey of Recent Research in Distributed Lags", (mimeo, Australian National University).

Hendry, D.F., A.R. Pagan and J.D. Sargan (1984), "Dynamic Specification" in *Handbook of Econometrics*, Z. Griliches and M. Intriligator (eds.) Vol. II, Amsterdam: North-Holland.

Hendry, D.F. and T. von Ungern-Sternberg (1979), "Liquidity and Inflation Effects on Consumers' Expenditure", in A.S. Deaton (ed.) Essays in the *Theory and Measurement of Consumers' Behaviour* (Cambridge University Press, Cambridge).

Hildreth, C. and J.P. Houck (1968), "Some Estimators for a Linear Model with Random Coefficients", *Journal of the American Statistical Association*, 63, 584-595.

Hinich, M.J. and D.M. Patterson (1985), "Identification of the Coefficients in a Non-Linear Time Series of the Quadratic Type", *Journal of Econometrics*, 30, 269-288.

Hoffman, D.L. and P. Schmidt (1981), "Testing the Restrictions Implied by the Rational Expectation Hypothesis", *Journal of Econometrics*, 15, 265-287.

Horowitz (1986), "A Distribution-Free Least Squares Estimator for Censored Linear Regression Models", *Journal of Econometrica*, 32, 59-84.

Huang, D.S. (1966), "The Short-Run Flows of Non-Farm Residential Mortgage Credit", *Econometrica*, 34, 433-459.

Ichimura, H. (1986), "Estimation of Single Index Models", (mimeo, M.I.T.).

Johnston, J. (1964), *Econometric Methods* (McGraw-Hill, New York).

Joiner, B.L. and J.R. Rosenblatt (1971), "Some Properties of the Range in Samples from Tukey's Symmetric Lambda Distributions", *Journal of American Statistical Association*, 66, 394-399.

Jonson, P.D., E.R. Moses and C.R. Wymer (1977), "The RBA 76 Model of the Australian Economy", in *Conference in Applied Economic Research* (Reserve Bank of Australia, Sydney).

Jorgenson, D.W. (1986), "Econometric Methods for Modelling Producer Behaviour", in Z. Griliches and M.D. Intriligator (eds.) *Handbood of Econometrics*, Vol. 3 (North-Holland, Amsterdam), 1841-1915.

Kalaba, R. and L. Tesfatsion (1985), "Nonlocal Comparative Statics, Automatic Derivatives, and Nonlinear Filtering: Recent Developments at USC", MRG Working Paper No. M8518, University of Southern California.

Kaplan, E.L. and P. Meier (1958), "Nonparametric Estimation from Incomplete Observations", *Journal of the American Statistical Association*, 53, 467-481.

Kennan, J. (1979), "The Estimation of Partial Adjustment Models with Rational Expectations", *Econometrica*, 47, 1441-56.

Kennan, J. (1985), "The Duration of Contract Strikes in U.S. Manufacturing", *Journal of Econometrics*, 28, 5-28.

Kenny, L.W., L.F. Lee, G. S. Maddala and R.P. Trost (1979), "Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data", *International Economic Review*, 20, 751-765.

Kennedy, P. (1985), A *Guide to Econometrics* (Blackwell, Oxford), 2nd edition.

Kiefer, N.M. (1985), "Specification Diagnotics Based on Laguerre Alternatives for Econometric Models of Duration", *Journal of Econometrics*, 28, 135-154.

Kiefer, J. and J. Wolfowitz (1956), "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters", *Annals of Mathematical Statistics*, 27, 887-906.

King, M.L. (1985), "A Point Optimal Test for Autoregressive Disturbances", *Journal of Econometrics*, 27, 21-37.

Koenker, R.W. (1982), "Robust Methods in Econometrics", *Econometric Reviews*, 1, 213-255.

Koenker, R.W. and G.W. Bassett (1978), "Regression Quantiles", *Econometrica*, 46, 33-50.

Koerts, J. and A.P.J. Abrahamse (1969), *On the Theory and Application of the General Linear Model* (Rotterdam University Press, Rotterdam).

Koopmans, T.C. (1950), "When Is an Equation System Complete for Statistical Purposes?", in T.C. Koopmans (ed.) *Statistical Inference in Dynamic Economic Models* (John Wiley, New York).

Kmenta, J. (1986), *Elements of Econometrics* (MacMillan, London), 2nd edition.

Kramer, W. (1982), "On the Consequences of Trend for Simultaneous Equation Estimation", *Economics Letters*, 14, 23-30.

Krasker, W.S. and R.E. Welsch (1982), "Efficient Bounded-Influence Regression Estimation", *Journal of the American Statistical Association*, 77, 595-604.

Kuh, E., J.W. Neese and P. Hollinger (1985), *Structural Sensitivity in Econometric Models* (John Wiley, New York).

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment", *Econometrica*, 47, 939-956.

Lancaster, T. (1985), "Generalized Residuals and Heterogeneous Duration Models", *Journal of Econometrics*, 28, 155-169.

Leamer, E. (1983), "Let's Take the Con out of Econometrics", *American Economic Review*, 73, 31-44.

Leamer, E. (1984), "Vector Autoregressions for Causal Inference", *Carnegie-Rochester Conference Series on Public Policy*, 22, 255-303.

Lee, L.F. (1979), "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables", *Econometrica*, 47, 977-996.

Lee, L.F. (1982), "Some Approaches to the Correction of Selectivity Bias", *Review of Economic Studies*, 49, 355-372.

Lee, L.F. and G.S. Maddala (1985), "The Common Structure of Tests for Selectivity Bias, Serial Correlation, Heteroscedasticity and Non-Normality in the Tobit Model", *International Economic Review*, 26, 1-20.

Lewis, H.G. (1974), "Comments on Selectivity Biases in Wage Comparisons", *Journal of Political Economy*, 82, 1145-1155.

Lucas, R.E. (1972) "Econometric Testing of the Natural Rate Hypothesis", in O.Eckstein (ed.), *The Econometrics of Price Determination Conference* (Board of Governors of the Federal Reserve System, Washington).

Lucas, R.E. (1976), "Econometric Policy Evaluation: A Critique", in *The Phillips Curve and Labour Markets* (Carnegie-Rochester Series on Public Policy No. 1, North-Holland, Amsterdam).

Lucas, R.E. and T.J. Sargent (1981), "After Keynesian Macroeconomics", in R.E. Lucas and T.J. Sargent (eds.) *Rational Expectations and Econometric Practice* (University of Minnesota Press), 295-320.

McAleer, M. and M.H. Pesaran (1986), "Statistical Inference in Non-Nested Econometric Models", *Applied Mathematics and Computation*,

McCallum, B. (1976), "Rational Expectations and the Natural Rate Hypothesis: Some Consistent Estimates", *Econometrica*, 44, 43-52.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior", in P. Zarembka (ed.) *Frontiers in Econometrics* (Academic Press, New York), 105-142.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice", in C. F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications* (M.I.T. Press, Cambridge, Mass.), 198-272.

McFadden, D.L. (1984), "Econometric Analysis of Qualitative Response Models", in Z. Griliches and M.D. Intriligator (eds.) *Handbook of Econometrics*, Vol. 2 (North-Holland, Amsterdam), 1395-1457.

McGuire, T.W., J.W. Farley, R.E. Lucas and R.L. Winston (1968), "Estimation and Inference for Linear Models in which Subsets of the Dependent Variable are Constrained", *Journal of the American Statistical Association*, 63, 1201-1213.

Maddala, G.S. (1971), "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data", *Econometrica*, 39, 341-358.

Maddala, G.S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics* (Cambridge University Press, Cambridge).

Maddala, G.S. (1986), "Disequilibrium, Self-Selection, and Switching Models", in Z. Griliches and M.D. Intriligator (eds.) *Handbook of Econometrics*, Vol 3, 1634-1688.

Manski, C.F. (1975), "The Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205-228.

Manski, C.F. (1984), "Adaptive Estimation of Non-Linear Regression Models", *Econometric Reviews*, 3, 145-194.

Manski, C.F. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics*, 27, 313-333.

Manski, C.F. and T.S. Thompson (1986), "Operational Characteristics of Maximum Score Estimation", *Journal of Econometrics*, 32, 85-108.

Manski, C.F. and S.R. Lerman (1977), "The Estimation of Choice Probabilities from Choice-Based Samples", *Econometrica*, 45, 1977-1988.

Manski, C.F. and D. McFadden (1981), "Alternative Estimators and Sample Designs for Discrete Choice Analysis", in C.F. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, (MIT Press, Cambridge, Mass.), 2-50.

Mizon, G.E. (1984), "The Encompassing Approach in Econometrics", in D.F. Hendry and K.F. Wallis (eds.) *Econometrics and Quantitative Economics* (Blackwell, Oxford).

Mizon, G.E. and J.F. Richard (1986), "The Encompassing Principle and its Application to Testing non-Nested Hypotheses", *Econometrica*, 54, 657-678.

Mundlak, Y. (1963), "Estimation of Production and Behavioral Functions From a Combination of Time Series and Cross Section Data", in C.F. Christ (ed.) *Measurement in Economics*, (Stanford University Press, Stanford).

Murphy, C.W., I.A. Bright, R.J. Brooker, W.D. Greeves and B.K. Taplin (1986), *A Macroeconometric Model of the Australian Economy for Medium-Term Policy Analysis* (Office of the Economic Planning Advisory Council, Canbera).

Muth, J.F. (1961), "Rational Expectations and the Theory of Price Movements", *Econometrica*, 29, 315-335.

Nadaraya, E.A. (1964), "On Estimating Regression", *Theory of Probability and Its Applications*, 9, 141-142.

Nelson, C.E. (1975), "Rational Expectations and the Estimation of Econometric Models", *International Economic Review*, 16, 555-561.

Nelson, C.R. and C. I. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, 10, 139-162.

Newey, W.K. (1984), "A Method of Moments Interpretation of Sequential Estimators", *Economics Letters*, 14, 201-206.

Newey, W.K. (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests", *Econometrica*, 53, 1047-1070.

Newey, W.K. (1986), "Efficient Estimation of Models with Conditional Moment Restrictions", (forthcoming, *Econometrica*).

Newey, W.K. (1987), "Adaptive Estimation of Regression Models Via Moment Restrictions", *Econometric Research Program Memorandum No.* 330, (Princeton University).

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, 55, 703-708.

Nicholls, D.F. and A.R. Pagan (1983), "Heteroscedasticity in Models with Lagged Dependent Variables", *Econometrica*, 51, 1233-1242.

Nickell, S. (1985), "Error Correction, Partial Adjustment and All That: An Expository Note", *Oxford Bulletin of Economics and Statistics*, 47, 119-30.

Olsen, R. (1982), "Distribution Tests for Selectivity Bias and a More Robust Likelihood Estimator", *International Economic Review*, 23, 223-240.

Orcutt, G. (1957), "A New Type of Socio-Economic System", *Review of Economics and Statistics*, 58, 773-797.

Orcutt, G., M. Greenberger, J. Korbel and A. Rivlin (1961), *Microanalysis of Socioeconomic Systems: A Simulation Study.* (Harper and Brothers, New York).

Orcutt, G. et al (1976), *Policy Exploration Through Microanalytic Simulation* (The Urban Institute, Washington).

Pagan, A.R. (1978), "Some Simple Tests for Nonlinear Time Series Models", *CORE Discussion Paper 7812*, Universite Catholique de Louvain.

Pagan, A.R. (1984), "Model Evaluation by Variable Addition", in D.F. Hendry and K.F. Wallis (eds.) *Econometrics and Quantitative Economics* (Blackwell, Oxford), 103-133.

Pagan, A.R. (1985), "Time Series Behaviour and Dynamic Specification", *Osford Bulletin of Economics and Statistics*, 47, 199-211.

Pagan, A.R. (1986), "Two Stage and Related Estimators and Their Applications", *Review of Economic Studies*, LIII, 517-538.

Pagan, A.R. and J. Shannon (1985), "Sensitivity Analysis for Linearised Computable General Equilibrium Models", in J. Piggott and J. Whalley (eds.) *New Developments in Applied General Equilibrium Analysis* (Cambridge University Press, Cambridge).

Pagan, A.R. and A. Ullah (1986), "The Econometric Analysis of Models Involving Risk Terms", *Centre for Economic Policy Discussion Paper No.* 127 (London).

Park, J. and P.C.B. Phillips (1986), "Statistical Inference in Regressions with Integrated Processes: Parts I and II", (mimeo, Yale University).

Patterson, K.D. and J. Ryding (1984), "Dynamic Time Series Models with Growth Effects Constrained to Zero". *Economic Journal*, 94, 137-43.

Pesaran, M.H. (1974), "On the General Problem of Model Selection", *Review of Economic Studies*, 41, 153-171.

Pesaran, M.H. (1981), "Identification of Rational Expectations Models", *Journal of Econometrics*, 16, 375-398.

Pesaran, M.H. (1986), "Econometrics", in *The New Palgrave*.

Pesaran, M.H. (1987), *The Limits to Rational Expectations*, (Blackwell, Oxford).

Phillips, P.C.B. (1982), "Exact Small Sample Theory in the Simultaneous Equations Model", in Z. Griliches and M.D. Intriligator (eds.) *Handbook of Econometrics*, Vol. 1, Ch. 8 (North-Holland, Amsterdam).

Phillips, P.C.B. (1987), "Time Series Regression with Unit Roots", *Econometrica*, 55, 277-302.

Phillips, P.C.B. and S.N. Durlauf (1985), "Multiple Time Series Regression with Integrated Processes", *Cowles Foundation Discussion Paper No. 768* (Yale University, New Haven).

Piggott, J. and J. Whalley (1985), *U.K. Policy and Applied General Equilibrium Analysis* (Cambridge University Press, Cambridge).

Plosser, C.I., G.W. Schwert and H. White (1982), "Differencing as a Test of Specification", *International Economic Review*, 23, 535-552.

Poirier, D.J. (1975), "Spline Lags: Why the Almon Lag Has Gone to Pieces", in D.J. Poirier, *The Econometrics of Structural Change* (North-Holland, Amsterdam).

Powell, A.A. (1969), "Aitken Estimators as a Tool in Allocating Predetermnined Aggregates", *Journal of the American Statistical Association*, 64, 913-922.

Powell, J.L. (1986), "Symmetrically Trimmed Lease Squares Estimation for Tobit Models", *Econometrica*, 54, 1435-1460.

Quandt, R.E. (1982), "Econometric Disequilibrium Models", *Econometric Reviews*, 1, 1-63.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis", *Journal of the Royal Statistical Society*, Series B, 31, 350-371.

Revankar, N.S. (1980), "Testing of the Rational Expectations Hypothesis", *Econometrica*, 48, 1347-1363.

Robinson, P.M. (1983), "Non-Parametric Estimators for Time Series", *Journal of Time Series Analysis*, 4, 185-207.

Robinson, P.M. (1986a), "Non-Parametric Methods in Specification", *Economic Journal*, Supplement, 96, 134-141.

Robinson, P.M. (1986b), "Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form", (forthcoming, *Econometrica*).

Rose, D.E. and J.G. Selody (1985), "The Structure of the Small Annual Model", *Technical Report No.* 40, Bank of Canada.

Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a Density Function", *Annals of Mathematical Statistics*, 27, 832-837.

Roy, A.D. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, 135-146.

Salant, S. (1977), "Search Theory and Duration Data: A Theory of Sorts", *Quarterly Journal of Economics*, 91, 39-57.

Salmon, M. (1982), "Error Correction Mechanisms", *Economic Journal*, 92, 615-629.

Sargan, J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.

Sargent, T.J. (1976), "A Classical Macroeconomic Model for the United States", *Journal of Political Economy*, 84, 207-237.

Sawa, T. (1978), "Information Criteria for Discriminating Among Alternative Regression Models", *Econometrica*, 46, 1273-1291.

Shiller, R.J. (1973), "A Distributed Lag Estimator Derived from Smoothness Priors", *Econometrica*, 41, 775-788.

Sims, C.A. (1972), "Money, Income and Causality", *American Economic Review*, 62, 540-552.

Sims, C.A. (1980), "Macroeconomics and Reality", *Econometrica*, 48, 1-48.

Sims, C.A., J.H. Stock and M. Watson (1986), "Inference in Linear Time Series Models with Some Unit Roots", (mimeo, Stanford University).

Singleton, J.J. (1985), "Testing Specifications of Economic Agents' Intertemporal Optimum Problems in the Presence of Alternative Models", *Journal of Econometrics*, 30, 391-413.

Small, K.D. and D. Brownstone (1982), "Efficient Estimation of Nested Logit Models: An Application to Trip Timing", Research Memorandum No. 296, Princeton University.

Stock, J.H. (1985), "Nonparametric Policy Analysis", (mimeo, Harvard University).

Stock, J. (1985), "Asymptotic Properties of Least Squares Estimators of Co-integrating Vectors", (mimeo, Harvard University).

Stoker, T.M. (1984), "Completeness, Distribution Restrictions and the Form of Aggregate Functions", *Econometrica*. 52, 887-908.

Stoker, T.M. (1986), "Aggregation, Efficiency and Cross-Section Regression", *Econometrica*, 54, 171-188.

Stone, C.J. (1975), "Adaptive Maximum Likelihood Estimation of a Location Parameter", *Annals of Statistics*, 3, 267-284.

Swamy, P.A.V.B. (1970), "Efficient Inference in a Random Coefficient Regression Model", *Econometrica*, 38, 311-323.

Tauchen, G. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models", *Journal of Econometrics*, 30, 415-443.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24-36.

Tuma, N. (1976), "Rewards, Resources and the Rate of Mobility: A Non-stationary Multivariate Stochastic Model", *American Sociological Review*, 41, 338-360.

Trivedi, P.K. (1985), "Distributed Lags, Aggregation and Compounding: Some Econometric Implications", *Review of Economic Studies*, L11, 19-36.

Ullah, A. (1987), "Nonparametric Estimation of Econometric Functionals", (mimeo, University of Western Ontario).

Wales, T.J. (1977), "On the Flexibility of Flexible Functional Forms: An Empirical Approach", *Journal of Econometrics*, 5, 183-193.

Wallis, K.F. (1972), "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations", *Econometrica*, 40, 617-636.

Wallis, K.F. (1980), "Econometric Implications of the Rational Expectations Hypothesis", *Econometrica*, 48, 49-73.

Watson, G.S. (1964), "Smooth Regression Analysis", *Sankhya Series* A, 26, 359-72.

White, H. (1980a), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.

White, H. (1980b), "Using Least Squares to Approximate Unknown Regression Functions", *International Economic Review*, 21, 149-170.

White, H. (1982a), "Maximum Likelihood Estimation of Misspecified Models", *Econometrica*, 50, 1-25.

White, H. (1982b), "Instrumental Variables Regression with Independent Observations", *Econometrica*, 50, 483-500.

Wickens, M.W. (1982), "The Efficient Estimation of Econometric Models with Rational Expectations", *Review of Economic Studies*, 49, 817-838.

Wickens, M.W. (1986), "The Estimation of Linear Models with Future Rational Expectations by Efficient and Instrumental Variable Methods", *Centre for Economic Policy Research* (London) *Discussion Paper No.111*.

Wu, D.M. (1973), "Alternative Tests of Independence Between Stochastic Regressors and Disturbances", *Econometrica*, 41, 733-750.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", *Journal of American Statistical Association*, 57, 348-368.

Zellner, A. (1969), "On the Aggregation Problem: A New Approach to a Troublesome Problem", in K.A. Fox, J.K. Sengupta and G.V.L. Narasimham (eds.), *Economic Models, Estimation and Risk Programming: Essays in Honor of Berhard Tintner*. (Springer-Verlarg, New York), 365-374.