LAW AND LIMITS OF ECONOMETRICS

By

Peter C.B. Phillips

February 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1397

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

LAWS AND LIMITS OF ECONOMETRICS *

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

19 July 2002

Abstract

We start by discussing some general weaknesses and limitations of the econometric approach. A template from sociology is used to formulate six laws that characterize mainstream activities of econometrics and the scientific limits of those activities.

Next, we discuss some proximity theorems that quantify by means of explicit bounds how close we can get to the generating mechanism of the data and the optimal forecasts of next period observations using a finite number of observations. The magnitude of the bound depends on the characteristics of the model and the trajectory of the observed data. The results show that trends are more elusive to model than stationary processes in the sense that the proximity bounds are larger. By contrast, the bounds are of smaller order for models that are unidentified or nearly unidentified, so that lack or near lack of identification may not be as fatal to the use of a model in practice as some recent results on inference suggest.

Finally, we look at one possible future of econometrics that involves the use of advanced econometric methods interactively by way of a web browser. With these methods users may access a suite of econometric methods and data sets online. They may also upload data to remote servers and by simple web browser selections initiate the implementation of advanced econometric software algorithms, returning the results online and by file and graphics downloads.

Key words and Phrases: Activities and limitations of econometrics, automated modeling, nearly unidentified models, nonstationarity, online econometrics, policy analysis, prediction, quantitative bounds, trends, unit roots, weak instruments.

JEL Classifications: C100, C500, C870

*Presented as the Sargan Lecture, Royal Economic Society, Warwick University, March 2002. Thanks go to the NSF for research support under Grant No. SES-00-92509.

A fundamental issue that bears on all practical economic analysis is the extent to which we can expect to understand economic phenomena by the process of developing a theory, taking observations and fitting a model. An especially relevant question in practice is whether there are limits on how well we can predict future observations using empirical models that are obtained by such processes. Not only are we interested in whether there are such limits, we also want to find some quantitative expression for them and to address the issue of whether these limits are attainable in practical empirical work. Forty years of empirical experience in macroeconomic forecasting suggests that there are limits to our capacity to make predictions about economic activity. In fact, the performance of aggregate predictions has improved little over this time in spite of much early optimism, enormous bodies of research in macroeconomic theory and modeling, improvements in econometric methods, and larger data sets of better quality.

A primary limitation on empirical knowledge is that the true model for any given data is unknown and, in all practical cases, unknowable. Even if a formulated model were correct it would still depend on parameters that need to be estimated from data. Often, data are scarce relative to the number of parameters that need to be estimated, and this is especially so in models that have some functional representation that necessitates the use of nonparametric or semiparametric methods. In such situations one might expect that the empirical limitations on modeling are greater than in finite parameter models. Second, all models are wrong. The models developed in economic theory are metaphors of reality, sometimes amounting to a very basic set of relations that are easily rejected by the data. Yet these models continue to be used, often because they contain a kernel of truth that is perceived as an underlying 'economic law'. Also, many see it as advantageous to use this information in crafting an empirical model even though it is at best only approximately true because to do so may well be better than using an entirely unrestricted system or an arbitrarily restricted one. Whether or not it is worthwhile doing so is, of course, an empirical matter.

Our discussion of these issues starts with the consideration of some maxims of econometrics that make explicit the activities and some of the weaknesses of the econometric approach. We formulate these in a light-hearted vein as 'laws of econometrics'. These laws of econometrics are not intended as universal truths. Instead, they purport to express the essence of what is being done in econometrics and to characterise some of the difficulties that the econometric approach encounters in explaining and predicting economic phenomena. The position we take in this discussion is related to views about modeling that have been suggested recently in Cartwright (1999) and Hoover (2001). Cartwright advances the notion that models can be interpreted as machines that generate laws (so-called nomological machines) and, even more flexibly, puts forward the view that the laws that may emerge from modeling are analogous to the morals that we draw from story-telling fables. Hoover (2001) takes a sympathetic but even more flexible position in arguing the case that economic modeling is useful to the extent that it sheds light on empirical relationships. As Hoover puts it, talking about formal laws seems to do nothing for economics –

“even accumulated falsifications or anomalies do not cause scientists to abandon an approach unless there is the prospect of a better approach on offer”, Hoover (2001, pp.54, 150). This position is similar to that of the Rissanen (1986, 1989) who argues against the concept of a true model and sees statistics as a “language for expressing the regular features of the data”.

Next, we discuss some proximity theorems that measure how close an empirical model can get (in terms of its likelihood ratio) to the true model in some parametric family. These theorems have been developed in joint work with Werner Ploberger (2001, 2002) and build on some earlier work in statistical theory by Rissanen (1986, 1987). The bounds in these proximity theorems depend on the data as well as on the model being used. A discovery in this research that is important in economic applications is that the magnitude of the bound depends on the presence and the nature of trends in the data. In particular, the bounds are greater for trending data than when the data are stationary, thereby giving quantitative expresssion to the intuitively appealing notion that trending data are harder to predict than data that do not trend. These theorems allow for finite parameter families and families with local misspecifications. Modeling algorithms then allow for gross misspecification within family groups. Proximity theorems for prediction are also provided in this approach, quantifying limits on empirical forecasting capability that are relevant in empirical work where specification is suspect. The present paper also discusses some cases of practical importance involving evaporating trends and nearly unidentified models. The latter have attracted recent interest in microeconometrics in applied models where only weak instruments are available for endogenous regressors (such as the use of the quarter of birth date as an instrument for schooling in earnings regressions, c.f. Angrist and Kruger, 1991).

We address some issues related to the possible attainment of these bounds in practical research and mechanisms for doing so. One mechanism that we consider involves the provision of econometric technology online using web servers that are accessible on a 24/7 basis. We describe a prototype web site that has been developed by the author to provide macroeconomic forecasts using automated model selection methods. Such web services offer one possible future for econometrics in which econometric methods are made available online to a wide range of consumers through the provision of automated modeling facilities. These facilities may involve resident databases that are made available to users or the option of uploading customer data for econometric analysis and forecast purposes.

The paper is organised into three parts. Part I postulates six law of econometrics. These laws provide a template for the discussion of the main activities of econometrics and what has distinguished the subject from other applications of statistical methods. The framework also offers us an opportunity to comment on recent lines of research and discuss the limitations of the econometric approach. The second part of the paper describes recent attempts to quantify the empirical limits of econometric methodology and of the forecasting capacity of empirical econometric models. The final part of the paper discusses the provision of econometric techniques as a web service, so that empirical econometric methods can be used by a wide range of possible consumers including practitioners, much as web users can presently view graphs of economic data like exchange rates or financial asset prices when they access financial sites online.

Part 1: Six Laws of Econometrics

The laws of econometrics that we propose below are not intended as universal scientific truths. Instead they are laws that characterise the activities and limitations of econometrics. These are serious issues. But it is useful to present them in a way that does not overstate our scientific contributions given the complexity of the real economic world. A self deprecating approach has the advantage that it often helps to pinpoint the essential limitations of a scientific approach to human economic behavior. In thinking about these matters I have found some useful related maxims that have been put forward in sociology.

A Template of Empirical Laws from Sociology

Paul Lazarsfeld, one of the fathers of modern mathematical sociology, founded the Bureau of Social Research at Columbia University in 1941 and established the field of mass media communications with a landmark study of the influence of the media on voting behavior (Lazarsfeld et al. 1944). In what is now folklore in the discipline, Lazarsfeld is credited with the enunciation of four laws of sociology. The laws were intended as a humorous summation of the limitations of the discipline. As far as I am aware, they have not before appeared in print[1]. I use them here as a template for suggesting some related laws of econometrics.

1: *Some Do, Some Don't* When all the modeling is done and the statistical analysis is complete, we are often left with the conclusion about human behavior that some individuals do certain things like buy a product, while others don't and our models simply do not explain it. These unaccounted aspects of human behavior represent heterogeneity across individuals. We are well acquainted with this need to allow for individual heterogeneity in modeling individual and firm behavior in microeconometrics.

2: *It's Different in the South* Similar to the heterogeneity we observe in individual behavior, we often find heterogeneity across regions. Lazarsfeld encapsulated this idea in the distinction between the two original regions of the USA - the Northern and Southern states. Here, the differences were so great in the nineteenth century that they precipitated civil war. The fact that there may be greater heterogeneity across regions than within regions needs to be taken into account in formulating models of human behavior and economic activity. At the same time, of course, there may well be greater cross section dependence within regions. Both factors affect the way we might formulate a realistic model.

[1] These laws of sociology were kindly communicated to me by Ronald Milavsky, a former student of Lazarsfeld.

3: *Hill People Always Cause Trouble* Social and economic life is not a level playing field. The reality is that some people (and some corporations) corner key resources and occupy the best real estate. With economic resources comes power and influence. With power and influence comes responsibility. Lazarsfeld translated economic inequality and the trappings of power and influence into trouble - trouble that sometimes only becomes manifest when it is discovered that the responsibilities that accompany power are not being met, as in ongoing investigations of the accounting practices of large corportations. The quantification of economic inequality and the study of its troubling as well as its beneficial efffects on society continue to be major concerns in both sociology and economics.

4: *Nothing Works in India* We all recognise the prospect that models fail and that sometimes they fail in a big way. Behavioral theories that are developed for one context or culture often founder completely in another. Diagnosing model failure is an issue that econometric modelers have confronted, and mechanisms for finding improved models that address the deficiencies of others have been developed. But there is also the prospect that the class of potentially useable models itself is so impoverished in relation to the generating mechanism that there is little prospect of improvement, and total model failure results. In such situations, most of our accepted paradigms of modeling provide little help and we are forced to turn to other alternatives. In economic forecasting, for example, when the models give results that are considered totally unrealistic, the modelers themselves make intercept adjustments to get the forecasts back on track, a practice that we will discuss with some analysis below.

The Laws of Econometrics

In a similar spirit, we now formulate six laws of econometrics. These laws encapsulate some of the features of the econometric approach, provide some practical maxims of applied econometrics, and point to some strengths and weakenesses of prevailing econometric methodology.

1: *Some Methods Work, Some Don't*

Econometrics has in large part been concerned with the development of statistical machinery that is appropriate for economic models and economic data. This developmental process ocurs because sometimes the usual statistical methods work well and sometimes they do not. The process is well illustrated by the steady progression of modeling practice and econometric methodology from the bivariate correlational studies of Fisher (1907) and others at the turn of the twentieth century, and the subsequent use of linear models in which the regressors were taken as fixed (e.g., Koopmans, 1937, Tinbergen, 1939), through to the development of simultaneous equations methodology in which the regressors may be jointly dependent or predetermined. The theory of identification, estimation and inference for simultaneous systems was the centrepiece of econometrics until at least the mid 1970's and involved major advances in econometric estimation, such as the systematic development of instrumental variable techniques by Sargan (1958, 1959). The following paragraphs briefly trace some features of this particular development drawing attention to some areas where the methodology is ongoing and showing that the knowledge boundary, where accepted practice falters, is never far away.

In recognition that simultaneous equations often suffer empirically from serially correlated errors (noted by Orcutt, 1949), an early direction in which the econometric methodology progressed was the accommodation of weakly dependent equation errors. The marriage of simultaneous equations and weak dependence, as Sargan (1959) called it, led to the development of new estimation procedures that worked better, at least asymptotically, such as generalized instrumental variables (GIVE) and later generalized method of moments (GMM) by Hansen (1982). In the last two decades, this work has further evolved into the large subject area of cointegration, which has succeeded in addressing three of the principal features of macroeconomic data - joint dependence, serial dependence and nonstationarity. The field is vast and has now reached a degree of maturity where we have efficient estimators based on semiparametric least squares (Phillips and Hansen, 1990) and parametric maximum likelihood (Johansen, 1988), and easily implemented test procedures and diagnostics. The purpose of all this research has been to produce new methods that work where conventional procedures fail. A large body of empirical evidence has now accumulated about the use of these procedures in practice, revealing that, while we have successfully produced a fairly complete theory of inference for unit root nonstationary and jointly dependent time series data, the linkage mechanisms between series often seem to be much more subtle than the linear concept of cointegration allows.

Recent research (inter alia, Jeganathan, 1998; Kim and Phillips, 1999; Robinson and Marinucci, 1999, 2001) has begun to tackle the difficulties of formulating and estimating models in which the time series are $I(d)$ and have memory that is characterised by a possibly fractional parameter d, thereby allowing for greater generality than when d is integer. The problems presented by these models of fractional cointegration seem considerably more complex than the $I(1)/I(0)$ (variable/error) case that is now common in applications. Both conceptual and technical issues are involved. Since the degree of nonstationarity (or memory) in the data (as well as the equation errors) is typically unknown, these parameters characterising memory need to be estimated in addition to any cointegrating relations among the variables. Furthermore, empirical evidence indicates that the degree of nonstationarity in economic data often differs significantly from one variable to another. For instance, interest rates, inflation, the money supply and income all appear to be nonstationary with individual memory parameters in the vicinity of 0.9, 0.6, 1.4, and 1.0, respectively (see Shimotsu and Phillips, 2002, for recent empirical estimates and valid confidence intervals based on an exact local Whittle estimator that consistently estimates the long memory parameter for any value of d). In consequence, no finite linear relation among these variables or subsets of them can be cointegrating in the conventional sense, even though it is very common to formulate empirical models that relate these variables in a linear way. Such relationships would, in fact, be unbalanced in terms of the memory characteristics of the data. Similar remarks apply to finite order vector autoregressions (VAR's) and structural VAR's, which have been in common use for many years as empirical models for these variables in applied macroeconomics. These linear models show us that present conceptualizations of cointegration and fractional cointegration do not allow the degree of flexibility needed to relate economic variables with differing memory characteristics and trend behavior, revealing some of the shortcomings of existing empirical methodologies. We are now beginning to understand the ways in which nonlinear transformations of nonstationary series affect stochastic order and memory properties (e.g., Park and Phillips, 1999), but we have not yet made significant headway on formulating relationships involving many variables with long memory characteristics. Modeling the stochastic relations between economic variables in a way that faithfully accommodates their differing individual memory characteristics as well as their apparent joint dependence is a task that exceeds present capability.

Another area of recent research on the boundary of our knowledge of simultaneous systems is the subject of estimation and inference with poor instruments. The early research of Koopmans and Hood (1953) and Sargan (1958) acknowledged concern about lack of identification and attempted to construct tests of underidentification that could shed light on the absence of identification and weak instrumentation. Much of this early research was neglected for many years. One exception is Sargan (1983) who devoted his Econometric Society presidential address to the consideration of models that were nearly unidentified, showing that slower convergence rates occured when first order rank conditions for identification failed but the parameters were still identified. Another is Phillips (1989), which provided a systematic study of the large sample properties of instrumental variables estimators for unidentified systems, showing that the estimates converged weakly to random variables that reflected the uncertainty about the parameters that was implicit in their lack of identification, and that Wald tests about unidentified parameters had the same distribution under both null and alternative hypotheses, foreshadowing later work on unbounded confidence intervals in such situations by Dufour (1997). The conclusion of Phillips (1989) argued the relevance of this new asymptotic theory to empirical work in microeconometrics where the low R^2 in companion regressions are often suggestive of weak instrumentation and near unidentification. Subsequent empirical work by Angrist and Kruger (1991) that related earnings to schooling using birth dates as instruments brought wide attention to this problem of weak instrumentation as an important issue for practitioners. A decade of theoretical research followed. Recent work in the field (Kleibergen, 2000; Forchini and Hillier, 2001; Moreira, 2001) seeks to make inference in weakly identified situations conditional on the amount of information in the data about the parameters on which identification hinges. We have also discovered that increasing the number of instruments can partially compensate for the fact that each of the instruments is weak. In fact, it is easy to show that one can obtain consistent estimates (but at reduced rates) as the number of instruments goes to infinity in situations where estimates with finite numbers of weak instruments converge weakly to random variables. All of this research makes it clear that, if the potential effects of weak instrumentation are not accounted for, inference can be badly distorted. Here again we find that the usual statistical methods do not work well, and conventional asymptotic properties can be a poor guide to the actual properties of estimation and testing procedures in practical situations.

2: *It's Different in Infinite Dimensional Spaces*

Much of modern econometrics is about trying to achieve generality wherever that is possible, but especially with regard to aspects of a model about which there is little prior knowledge. On the other hand, where a model connects most closely with some underlying economic hypothesis, we often seek to retain specificity through direct parameterisation. These considerations have led to a flowering of work in the last two decades on nonparametric and semiparametric estimation. These methods are now used in financial econometrics (recent examples being the functional estimation of diffusion equations - see Bandi and Phillips, 2002), time series (a major field of recent application being that of the semiparametric estimation of the long memory parameter in an $I(d)$ process - see Baillie, 1996, and Henry and Zaffaroni, 2001, for reviews), and microeconometrics (where adaptation for heterogeneity of unknown form or unknown error distributions is often important - see Horowitz, 1998, for an overview).

Estimating a function with a finite amount of data is like running a marathon. A marathon is not run in a series of 100 metre sprints. Instead, the available energy (or data) is spread out so that it lasts for the whole course and contributes to estimation over the full domain of the function. In consequence, one has slower rates of convergence in function space, typically at a $\sqrt{nh}$ rather than $\sqrt{n}$ rate, where n is the number of data points and h is a bandwidth parameter that controls the width of the band used around each local point in the domain. Complications arise over the choice of bandwidth and the fact that the data is spread more thinly when the dimension of the function space increases, leading to commensurate reductions in the convergence rate. Because the rate of convergence is slower, asymptotic theory is often a less satisfactory device for producing adequate approximations to the distribution of function estimates. Asymptotic expansions are especially helpful here because they can offer improvements on first order limit theory and quantitative insights that can guide suitable bandwidth choices (Linton, 1996; Xiao and Phillips, 1998, 2002). On the other hand, these expansions are more complex than conventional expansions because they involve the two parameters n and h. Function estimation can also be used when the data is nonstationary, either to estimate the amount of time spent by the process in various spatial vicinities (see Phillips, 1998, 2001) or to provide nonparametric estimates of drift and diffusion functions in potentially nonstationary diffusion equations (Bandi and Phillips, 2002). In these and other respects, both theory and application are different in infinite dimensional space.

3: *Unit Roots Always Cause Trouble*

Unit roots are the new hill people of econometrics. Unless you are a Bayesian (Sims and Uhlig, 1991; Kim, 1994; Phillips and Ploberger, 1996), unit roots inevitably cause trouble because of the nonstandard limit distributions (see Phillips and Xiao, 1998, for a recent review) and the discontinuities that arise in the limit theory as the autoregressive parameter passes through unity (but see Phillips, 1987b, Chan and Wei, 1987, and Phillips, Moon and Xiao, 2001, for attempts to unify this asymptotic theory). The nonstandard limit distributions themselves vary, depending on the specification of the model and any prior filtering (such as demeaning or deterministic detrending) that has been done in the estimation of the autoregressive coefficient (Park and Phillips, 1988, 1989). So, the commonplace filtering and regression with integrated time series that is done in the econometric kitchen inevitably shows up in the attic in the asymptotic theory. The situation is analogous to that of the fictional character Dorian Gray in the novel by Oscar Wilde (1890) - the face of Dorian Gray showed no signs of aging as time passed, whereas the sins of his worldly existence showed up to torment him in the portrait of himself that he hept hidden in the attic.

Unit roots also cause trouble because of the difficulty in discriminating between stochastic trends and deterministic trend alternatives, including models that may have trend breaks. Much of the received wisdom on this subject focuses on what is perceived as the poor power properties of unit root tests and stationarity tests. However, an alternate perspective is that unit roots and deterministic trending processes may both have validity in explaining the same characteristics of the data, viz. their trending behaviour. With this perspective, the issue subtly changes from the adversarial position of stochastic trends versus trend stationarity to one in which many competing explanations are admitted as possible. Econometric practice can then focus on finding those models and explanations that are the most useful. We describe this alternate perspective more fully below. It is an interesting feature of the research process that, in spite of the enormous amount of work that has been done on unit root theory and testing in the last two decades, subtler issues such as these are only now being considered.

In some respects, panel unit root problems cause even more trouble. In the first place, the asymptotic theory is often multidimensional with both the cross section sample size (N) and the time series sample size (T) tending to infinity. Situations of this type are studied in Phillips and Moon (1999) and, depending on the passage of N and T to infinity, we can get both standard and nonstandard limit theory. In cases where T is fixed, bias problems in dynamic panel estimation is known to be severe and to lead to inconsistencies in estimation by maximum likelihood (e.g., Nickell, 1982). Bias is further aggravated in the unit root case (Phillips and Sul, 2001) and even occurs when both N and T tend to infinity in the case of near unit roots (Moon and Phillips, 1999, 2001). These are all instances of the incidental parameters problem (e.g., Lancaster, 1998) that arises when there is a proliferation of nuisance parameters from fixed effects and individual specific trends. In such cases, maximum likelihood, in attempting to get good estimates of all the parameters in a model, ends up failing to obtain consistent estimates of some them. In dynamic panel models, the inconsistency unfortunately shows up in the important autoregressive coefficient that governs the dynamics.

As Maddala (Lahiri, 1999) remarks, much of the original interest in the problem of panel unit roots was to assess whether there was homogeneity in dynamic behaviour across individuals in the panel – the question, in effect, was whether unit roots really persisted across individuals in a panel. Homogeneity testing of this type remains an extremely important issue in practical work. By contrast, much of the attention in theoretical work has focused on the gains to be had from pooling cross section observations under homogeneity. In unit root cases, the gains can appear substantial because pooling converts nonstandard into more useable standard limit theory by cross section averaging. These features have made panel unit root theory popular among practitioners. How relevant these results are when homogeneity does not apply is a different matter. Another issue is how well the pooled limit theory holds up when the asymptotics are multidimensional and $T \to \infty$ more slowly than $N \to \infty$. Many of these practically important matters need investigation. The field is vast and there is a lot to be done.

4: *Cross Section Dependence also Causes Trouble*

It is convenient and has for long been common econometric practice to assume cross section independence in panel modeling up to a time specific effect. Yet cross section dependence is often to be expected in microeconomic applications of firm and individual behaviour, is almost always present in regional or cross country macroeconomic panels, and it is widely acknowledged as a major characteristic of financial panels.

In recognition of its empirical relevance, cross section dependence is a rapidly growing field of study in panel data analysis. But there are many limitations to the models being used and unresolved difficulties for empirical workers. A primary difficulty arises because there is no natural ordering of cross section data, making it hard to characterise and model dependence across section. This difficulty is exacerbated by the absence of a theory justifying (or even suggesting) realistic forms of cross section dependence. Without theory, there are few restrictions on the degree of dependence that can be imposed *a priori*. Increases in cross section sample size then lead to a rapid proliferation in the number of parameters to be estimated and potential incidental parameter problems like the inconsistency problem mentioned above. One approach in dynamic modeling has been to use a factor structure and delimit the number of factors to one or two (Phillips and Sul, 2002) or use model selection methods to empirically determine the number of factors (Bai and Ng, 2002; Moon and Perron, 2001). Once a factor structure is determined, the estimation of dynamic factors presents further difficulties. The obvious approach here is to use principal components (Stock and Watson, 1998, 1999; Bai and Ng, 2001, Moon and Perron, 2001). But moment based approaches (Phillips and Sul, 2002) seem to offer an interesting alternative. A fully fledged asymptotic theory for $(N, T) \to \infty$ is still to be developed, and assessment of the various alternative approaches is hampered by the many different ways in which the nuisance parameters can be treated and the absence of an optimal theory of estimation. Traditional asymptotic theory for panels (see Maddala, 1993, for an overview) conditioned on a fixed value of T, typically assumed time series stationarity, relied heavily on cross section independence, and involved the passage to infinity of only the single index N. Panel data research in the last decade has begun to address each of these issues, but awaits a systematic multi-index asymptotic analysis that allows for cross section dependence and general forms of time series nonstationarity.

5: ***No One Understands Trends.***

In spite of the importance of trends in macroeconomic research, particularly in the study of economic growth and growth convergence, economic theory provides little guidance for empirical research on the formulation of trend functions. This partly explains the rather impoverished class of trend formulations that are in use in econometrics. Most commonly, these are polynomial time trends, simple trend break polynomials, and stochastic trends, which include unit root models, near unit root models and fractional processes. More occasionally, sinusoidal time polynomials and nonparametric trend specifications are used. When the focus is on trend elimination (for instance, in the extraction of the cyclical component of a series for studying business cycles), smoothing methods are popular. The most prominent of these is the Whittaker (1923) filter, which is commonly known in macroeconomics as the Hodrick-Prescott (1980) filter, and the closely related spline smoothers (Schoenberg, 1964; Wahba, 1978). Band pass filters like those in Baxter and King (1999) and Corbae, Ouliaris and Phillips (2002) are also used. All these methods provide a mechanism for dealing with trends in the data. But it is unrealistic to pretend that such formulations and filters explain the process by which trends actually occur in the real world. In short, no one really understands trends, even though most of us see trends when we look at economic data.

One nearly universal consequence of trends in the data is the regression phenomena called spurious regression. In effect, any trend function that we postulate in an econometric specification will turn out to be statistically significant in large samples provided the data do in fact have a trend, whether it is of the same form as that specified in the empirical regression or not. Perhaps the most well known example is that polynomial trends are statistically significant (with probability one, asymptotically) when the true trend is stochastic and vice-versa (Durlauf and Phillips, 1988). This is so even when robust standard errors are used to assess significance (Phillips, 1998). Similar results hold for trend breaks, fractional processes and regressions among such variables even when they are stochastically independent, the phenomenon originally studied in Granger and Newbold (1974) and Phillips (1986).

The nomenclature 'spurious regression' has become universal and carries a pejorative connotation that generally makes empirical researchers anxious to show that their fitted relationships are validated by some procedure such as a test for cointegration. An alternative perspective proposed in Phillips (1998) is that deterministic trend functions (or even the time path of another trending variable) can be used as a coordinate system for measuring the trend behavior of an observed variable, much as one set of functions can be used as a coordinate basis for studying another function. For instance, we can write any function $f \in L_2[0,1]$ in terms of an orthonormal basis $\{\varphi_k\}_{k=1}^{\infty}$ as $f(x) = \sum_{k=1}^{\infty} c_k \varphi_k(x)$. Continuous stochastic processes such as Brownian motion and diffusions also have representations in terms of the functions φ_k but with coefficients c_k that are random variables rather than constant Fourier coefficients. In a similar way, we can write trending data in terms of coordinates comprised of other trends, like time polynomials, random walks or other observed trends. Such formulations can be given a rigorous function space interpretation in terms of functional representations of the limiting stochastic processes or deterministic functions to which standardized versions of the trending data or trend functions converge. What is particularly interesting about this perspective is that it provides a mechanism for relating variables of different stochastic order (like time polynomials and random walks) so that it can be used to justify relationships between observed variables like interest rates, inflation, money stock and GDP, which have differing memory characteristics, overcoming the problem of stochastically imbalanced relationships discussed earlier. This approach also offers an interpretation of empirical regressions that are deliberately constructed to be spurious such as the celebrated example of prices on cumulative rainfall (Hendry, 1980). Here, cumulative rainfall is a stochastic trend by construction and this trend is simply one possible coordinate (by no means a good one *a priori*) for measuring the trending behavior of prices. Of course, other coordinates, like the aggregate stock of money, may well provide a more economically meaningful coordinate system, but this does not invalidate the rainfall aggregate as a potential yardstick for assessing the trend in price levels.

A secondary element in this alternative perspective of spurious regression is that when we include different trend functions in an empirical regression, they will each compete in the explanation of the observed trend in the data. Correspondingly, when we regress a unit root stochastic trend on a time polynomial of degree K as well as a lagged variable, each of the $K+1$ regressors is a valid yardstick for the trend. If we let $K \to \infty$ as the number of observations $n \to \infty$ but with $\frac{K}{n} \to 0$ so that the regression remains meaningful as n grows large, then the coefficient of the lagged variable tends to unity but at the reduced rate $\frac{n}{K}$. This reduction in the rate of convergence to a unit root coefficient demonstrates how seemingly irrelevant time polynomial regressors can succeed in reducing the explanatory power of a lagged dependent variable even though the true model is a first order autoregression (Phillips, 2002).

The previous discussion speaks to the importance of misspecification analysis in studying trends. Recognising that trend specifications are inevitably wrong in empirical practice has implications for forecasting. The subject has received little attention in the literature, with the recent exception of Clements and Hendry (1999, 2001). The following brief analysis gives some new results, showing how we can still perform useful forecasting exercises despite the presence of (inevitably) misspecified trends.

Suppose that X_t is a stochastic trend with $\Delta X_t = u_t$ and that partial sums of the stationary process u_t satisfy the functional law $n^{-1/2} \sum_{k=0}^{[n\cdot]} u_t \to_d B(\cdot)$, a limit Brownian motion process. Suppose also that X_t is erroneously modeled by a linear deterministic trend, giving the (spurious) regression equation $X_t = \hat{b}_n t + \hat{u}_t$, where $\hat{b}_n = \sum_{t=1}^{n} X_t t / \sum_{t=1}^{n} t^2$. It is often suggested that the behavior of forecasts from such erroneous regressions is one of the more serious consequences of misspecification. The h - period projection of the fitted linear trend, $\hat{b}_n(n+h)$, for instance, seems very different from the constant level prediction X_n one gets from a martingale model for X_t. In fact, $\hat{b}_n(n+h)$ produces divergent behavior as h becomes large. However, the situation is more benign than appears to be generally known. For instance, one period ahead forecasts from the fitted trend have the form $\hat{X}_{n+1} = \hat{b}_n\,(n+1)$, and since (Durlauf and Phillips, 1988) $\sqrt{n}\hat{b} \to_d \int_0^1 rB / \int_0^1 r^2$, it follows that

$$\hat{b}_n = O_p\left(\frac{1}{\sqrt{n}}\frac{\int_0^1 rB}{\int_0^1 r^2}\right) = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{1}$$

so that

$$\hat{X}_{n+1} = O_p\left(\sqrt{n}\right),$$

which is precisely the same stochastic order as the optimal forecast

$$\tilde{X}_{n+1} = X_n + E_n\left(u_{n+1}\right) = O_p\left(\sqrt{n}\right). \tag{2}$$

Moreover, if intercept correction using the last period error (see Clements and Hendry, 1999, for a recent discussion) is employed, the following adjusted forecast is obtained

$$\check{X}_{n+1} = \hat{X}_{n+1} + (X_n - \hat{X}_n).$$

Direct calculation reveals that

$$\begin{aligned}
\check{X}_{n+1} &= \hat{b}_n\,(n+1) + \left(X_n - \hat{b}_{n-1}n\right) \\
&= X_n + \hat{b}_n + \left(\hat{b}_n - \hat{b}_{n-1}\right)n \\
&= X_n + O_p\left(\frac{1}{\sqrt{n}}\frac{\int_0^1 rB}{\int_0^1 r^2}\right) + O_p\left(\frac{n^2 X_n \sum_{t=1}^{n-1} t^2 - n^3 \sum_{t=1}^{n-1} X_t t}{\sum_{t=1}^{n-1} t^2 \sum_{t=1}^{n} t^2}\right) \\
&= X_n + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Thus,

$$\check{X}_{n+1} = \tilde{X}_{n+1} - E_n\left(u_{n+1}\right) + O_p\left(\frac{1}{\sqrt{n}}\right), \tag{3}$$

so that the intercept adjusted forecast from the misspecified model, $\check{X}_{n+1}$, differs from the optimal forecast, $\tilde{X}_{n+1}$, by the stationary process $E_n\left(u_{n+1}\right)$ up to an error of $O_p(n^{-1/2})$. Thus, prediction from a misspecified trend may not be that serious provided we make an effort to keep the model on track by using intercept adjustments. This is, of course, a time-honored empirical practice in applied forecasting (e.g., Evans, 2002).

In fact, we can go further than this. Take the observed prediction errors of the adjusted forecasts giving

$$\begin{aligned} X_{n+1} - \check{X}_{n+1} &= X_n + u_{n+1} - \check{X}_{n+1} \\ &= u_{n+1} + O_p\left(\frac{1}{\sqrt{n}}\right) \\ &= E_n\left(u_{n+1}\right) + \varepsilon_{n+1} + O_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{4}$$

where $\varepsilon_{n+1} = u_{n+1} - E_n\left(u_{n+1}\right)$ is a martingale difference. The prediction errors $X_{n+1} - \check{X}_{n+1}$ therefore differ from the original stationary residual process u_{n+1} by a term of $O_p(n^{-1/2})$; and they are asymptotically the same as $\tilde{X}_{n+1} - \check{X}_{n+1}$, or $E_n\left(u_{n+1}\right)$, up to a martingale difference. We may then model the prediction errors $X_{n+1} - \check{X}_{n+1}$ using stationary process techniques to obtain an empirical estimate, $\hat{E}_n$ say, of the stationary sequence $E_n\left(u_{n+1}\right)$. Using this estimate, we can modify the adjusted forecasts $\check{X}_{n+1}$ to construct a predictor that is fully adjusted for specification errors in the trend and stationary components, viz.

$$\check{X}^{+}_{n+1} = \check{X}_{n+1} + \hat{E}_n. \tag{5}$$

From (4) and (5), it is apparent that

$$\begin{aligned} X_{n+1} - \check{X}^{+}_{n+1} &= u_{n+1} - \hat{E}_n + O_p\left(\frac{1}{\sqrt{n}}\right) \\ &= \varepsilon_{n+1} + \left(E_n\left(u_{n+1}\right) - \hat{E}_n\right) + O_p\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

which differs from the optimal forecast error $\varepsilon_{n+1} = X_{n+1} - \tilde{X}_{n+1}$ by the error of stationary estimation $E_n\left(u_{n+1}\right) - \hat{E}_n$, and a term of order $O_p(n^{-1/2})$. In this way, intercept adjustment compensates for trend misspecification and enables subsequent modification to achieve forecasts that are asymptotically equivalent to the optimal forecast $\tilde{X}_{n+1}$.

Similar results can be shown to hold in the alternate case where the true model is a trend stationary process and the supposed model is a unit root stochastic trend. It may be true that no one understands trends. But if we acknowledge the inevitable presence of trend misspecification and adjust forecasts often and with enough care, then we may be able to make do with our existing impoverished arsenal of trend specifications.

6: *DGP's are Unknown and Inherently Unknowable.*

Having collected data and knowing that the process by which it has been gathered can be well described, it seems like a simple step to accept the notion that there must be a corresponding 'true model' or data generating process (dgp). However, whether that process can be faithfully and completely represented in terms of a formal statistical model whose variables are defined on a certain probability space is a different matter altogether. It often seems reasonable to think of quantifiable economic variables as random variables defined on a probability space and this has proved to be a very useful practical approach to formal modeling. Indeed, we deal so much with models, random processes and probability spaces in our work as econometricians that is easy to be lured into thinking that there must be an underlying true dgp. However, the actual process of data generation may not fit faithfully into this framework without an extraordinary level of additional complexity that belies the notion of modeling as we presently know it. This view may at first appear heretical but it becomes more reasonable upon reflection. In the case of econometric modeling, we may note that individual decision makers rarely make purely random decisions, much less ones that follow nice Gaussian distributions, and the factors that enter decision making are often so numerous and complex as well as individual specific that it is hard to conceive of a probability space large enough to capture all of the determining factors accurately. A series such as national income illustrates many of the essential problems. The process by which individual incomes are aggregated into national income has been carefully defined according to certain conventions and we may reasonably take each component income in the aggregate to be a quantifiable random variable. Yet, individual income components are determined in differing ways that can be extremely complex, depending as they often do on a host of decisions made at different points in time by different personnel involved in the hiring, promotion and wage determination process, as well as a vast number of historical, institutional, regional and local precedents that bear on wage setting behavior. To faithfully capture all of these elements, as well as their many endogeneities and dependencies, in a true model of individual specific wage determination for each of the individuals in the aggregate seems like an impossible task. Indeed, to do so is antithetic to the notion that a model itself is a simplified representation of a real world process. Correspondingly, any attempt to faithfully represent a variable like national income in terms of a formal statistical model like an autoregression or an autoregression with distributed lag effects from other variables is a heroic simplification where the distribution of the error component only crudely captures the omitted influences. Similar comments apply to more general attempts at modeling, such as nonparametric approaches.

The ideas about modeling and true dgp's expressed in the last paragraph have many antecedents. Hannan (1993), for instance, put the notion quite simply by saying that there is "never an attainable true system generating the data" and that the best that can be hoped for is that "a very restricted model class can be successfully used". Rissanen (1987) expressed similar views when he characterised statistical modeling as "a language to express the regular features of the data". These positions make sense as more realistic representions of the goals of statistical modeling of observed data than the idea of searching for or approximating an underlying true dgp. They are also highly suggestive of the notion discussed in the Introduction that there are limits to empirical knowledge.

If there were a true statistical model responsible for generating the observed data, such a model would inevitably involve elements unknown to an empirical researcher, such as the functional form of the systematic component of the model, the distribution of random error components or the true values of parameters or hyperparameters in the system. The simplest case would be one in which only the true values of the parameters were unknown and everything else from the model class to functional form were known and correctly specified. This obviously represents an ideal situation. Even in such a case the true dgp is still unknown. However, it is interesting to ask how close to the true system we can get using observed data in this ideal situation. As it turns out, there is a quantifiable bound on how close we can get to the true system and how well we can predict using an empirical model, which we now move on to discuss.

Part 2: Quantifying the Limits to Empirical Knowledge

In this ideal situation where there is a true system and the only unknowns are a finite number of parameters to be estimated, closeness to the true system depends on how well we can estimate these parameters and the role these parameters play in generating the data. Our discussion here will focus here on the time series case. We will briefly report some proximity theorems of Ploberger and Phillips (2002) that deliver quantitative bounds on how close empirical models can come to the true system in this context and discuss some extensions of that theory.

Proximity Bounds in Modeling and Forecasting[2]

The line of reasoning used in this research was pioneered by Rissanen (1986, 1987, 1996). Rissanen asked how close on average (measured in terms of Kullback-Leibler (KL) distance) can we get to a true dgp using observed data. It is presumed that time series data $\{X_t\}_{t=1}^n$ is available and the dgp belongs to a k-dimensional parametric family and satisfies certain regularity conditions. The dgp is known up to a certain parameter θ and P_θ is the corresponding unknown probability measure. The class of potential empirical models for the data generated by P_θ is very wide, but will normally depend on some rules of estimation for obtaining numerical values of the unknown parameters or rules for averaging the parameters out, both leading to a usable empirical form of the model that can be represented by a proper probability measure, G_n, say. The most common empirical models are constructed using classical and Bayesian principles. In the classical approach (or in Dawid's, 1983, terminology, the prequential approach) unknown parameters are replaced by their maximum likelihood estimates, whereas in the Bayesian approach the unknown parameters are averaged out to produce the data density or marginal likelihood.

As a measure of 'goodness of fit' to the true model we may use the sequence of random variables given by the log likelihood ratio

$$\ell_n(G_n) = \log \frac{dG_n}{dP_\theta},$$

computed for different empirical models G_n. Rissanen (1987,1996) showed that if X_t is stationary, if $\theta \in \Theta$, a regular subset of $\mathbb{R}^k$ (i.e. $\dim \Theta = k$), and if some technical

[2] The discussion in this section draws on Ploberger and Phillips (2001, 2002).

conditions are fulfilled, then the Lebesgue measure (i.e., the volume in $\mathbb{R}^k$) of the set

$$\left\{\theta : -E_\theta \log \frac{dG_n}{dP_\theta} \leq \frac{1}{2} k \log n\right\} \tag{6}$$

converges to 0 as $n \to \infty$ for any choice of empirical model G_n. This theorem shows that, whatever one's model, one can approximate (with respect to KL distance) the dgp no better on average than $\frac{1}{2}k \log n$. Thus, outside of a 'small' set of parameters we can get no closer to the truth than the bound $\frac{1}{2}k \log n$, and the 'volume' of the set for which we can do better actually converges to zero.

Rissanen's theorem justifies a certain amount of skepticism about models with a large number of parameters. The minimum achievable distance of an empirical model to the dgp in this theory increases linearly with the number of parameters. In essence, the more complex the system is, the harder it is to construct a good empirical model. Thus, the theorem makes precise the intuitive notion that complex systems can be very hard to model, that models of larger dimension place increasing demands on the available data. The bound $\frac{1}{2}k \log n$ in (6) provides a yardstick for how 'close' to the true probability measure we can get within a parametric family, assuming that the parameters all have to be estimated with the given data. An important feature of this Rissanen bound is that it treats all parameters equally by way of the fact that it depends on the total number of parameters k.

Ploberger and Phillips (2002) pursue a similar analysis but gave almost sure (rather than average) proximity results and worked with a broader class of assumptions that allow for some nonstationary as well as stationary time series. They gave a general 'limitation result' for regressions with integrated and cointegrated variables as well as stationary time series, and validated the higher level assumptions of the theory for simultaneous equations models. In their result, an important role is played by the conditional 'Fisher information' matrix, $B_n = \sum_{1\leq i \leq n} E_\theta(\varepsilon_i(\theta)\varepsilon_i(\theta)'|\mathfrak{F}_{i-1})$, where $\varepsilon_i(\theta) = \frac{\partial}{\partial\theta} \log p_\theta(X_i|\mathfrak{F}_{i-1})$ is a score component and $p_\theta(X_i|\mathfrak{F}_{i-1})$ is the conditional density corresponging to $P_\theta(\cdot|\mathfrak{F}_{i-1})$ and where $\mathfrak{F}_i$ is a filtration. Ploberger and Phillips show that for any empirical model G_n and every compact set K in the parameter space, the Lebesgue measure of the set of structures

$$\left\{\theta : P_\theta\left(\left[-\log\left(\frac{dG_n}{dP_\theta}\right) \leq \frac{1-\varepsilon}{2} \log \det B_n(\theta)\right]\right) \geq \alpha\right\} \cap K \tag{7}$$

converges to 0 as $n \to \infty$, for any given small $\varepsilon > 0$ and some $\alpha > 0$. This result means that sets of θ for which the empirical model G_n can do better than the bound $\frac{1-\varepsilon}{2} \log \det B_n(\theta)$ with nonneglible probability $\alpha > 0$ have volume in $\mathbb{R}^k$ that goes to zero as the sample size $n \to \infty$. In other words, up to a small exceptional set in $\theta-$space, no empirical model G_n can come closer to the true dgp than $\frac{1-\varepsilon}{2} \log \det B_n$, a bound that depends on the data through B_n. The bound may well therefore be path dependent, rather than being reliant solely on the dimension of the parameter space, and there is no reason why it will treat parameters equally. Indeed, coefficients of trending regressors actually increase the bound even though these coefficients may be estimable at higher rates than the coefficients of stationary variables.

Most of the commonly arising cases in time series econometrics lead to asymptotic expressions of the form

$$\log \det B_n \sim \left(\sum_{i=1}^{k} \alpha_i \right) \log n \tag{8}$$

for the sample information where $\alpha_i \geqq 1$ with inequality occuring for at least one element i when there are trending mechanisms in the model. In particular, $\alpha_i = 1$ for stationary regressors, $\alpha_i = 2$ for stochastic trends, $\alpha_i = 2d$ for regressors with long memory d, and $\alpha_i = 3$ for a linear deterministic trend. These scale coefficients α_i make it clear that the achievable distance of an empirical model to the dgp increases faster the stronger is the trending behavior. In effect, when nonstationary regressors are present, it appears to be even more important to keep the model as simple as possible. In particular, an additional stochastic trend in a linear regression model will be twice as expensive as a stationary regressor in terms of the marginal increase in the nearest possible distance to the dgp and a linear trend three times more expensive. Although nonstationary regressors embody a powerful signal and have estimated coefficients that display faster rates of convergence than those of stationary regressors, they can also be powerfully wrong in prediction when inappropriate and so the loss from including nonstationary regressors is correspondingly higher. One of the conclusions of this work, therefore, is that in a clearly quantifiable sense the true dgp turns out to be more elusive when there is nonstationarity in the data.

The above results apply irrespective of the modelling methodology that is involved. Neither Bayesian nor classical techniques nor other methodologies can overcome this bound on empirical modelling. The bound can be improved only in 'special' situations – special because the sets for which improvements can occur have Lebesgue measure zero in $\mathbb{R}^k$ – like those where we have extra information about the true dgp and do not have to estimate all the parameters. For instance, we may 'know' that there is a unit root in the model, or by divine inspiration we may hit upon the right value of a parameter and decide not to estimate it.

Result (7) has a counterpart in terms of the capacity of an empirical model to capture the good properties of the optimal predictor (i.e. the infeasible predictor that uses knowledge of the dgp and, in particular, the values of its parameters). Ploberger and Phillips (2002) show that for a general class of Gaussian simultaneous equation models, the limitations of an empirical model such as G_n in (7) carry over to the weighted forecast mean square divergence

$$\Delta_n = \sum_{t=n_0}^{n} \{ (y_t - \widehat{y}_t)' \Sigma^{-1} (y_t - \widehat{y}_t) - (y_t - \widetilde{y}_t)' \Sigma^{-1} (y_t - \widetilde{y}_t) \}, \tag{9}$$

where n_0 is some point of initialization for the optimal (one period ahead) forecasts $\widetilde{y}_t$ and another predictor for $\widehat{y}_t$, say, which is constructed from G_n and is $\mathfrak{F}_{t-1}$-measurable. In particular, there exists a number $A = \sum_{i=1}^{k} \alpha_i$ (depending on the degree of nonstationarity and taking into account cointegrating rank) which has the property that for Lebesgue almost all parameters and for all $\varepsilon > 0$

$$P_\theta\left(\left[\Delta_n \leq \frac{1-\varepsilon}{2} A \log n\right]\right) \to 0. \tag{10}$$

Thus, only on exceptional $\theta-$ sets can we expect to come closer (in terms of the divergence measure Δ_n) to the optimal forecast than the bound $\frac{1-\varepsilon}{2} A \log n$ as $n \to \infty$. So, in cases where the data is nonstationary, something new happens in prediction. Our capacity to get near to the optimal predictor is reduced whenever we include a nonstationary regressor. In the rule for determining empirical limits, we have to multiply the number of parameters by an additional factor that is essentially determined by the number and type of the trends in the regressors. Increasing the dimension of the parameter space therefore carries a price in terms of the quantitative bound of how close we can come to replicating the optimal predictor. This price goes up when we have trending data and when we use trending regressors.

What happens under Weak Identification?

In contrast to the case of trending regressors, the price of including additional regressors goes down when the signal diminishes, as it often does in cases of weak identification. For example, in the evaporating trend model

$$y_t = \frac{\beta}{t^\alpha} + u_t, \quad t = 1, ..., n; \quad u_t \sim iid\ N(0, \sigma^2), \quad \alpha = \frac{1}{2}, \tag{11}$$

we have $B_n = \sum_{t=1}^n \frac{1}{t} = \log n + O(1)$, and so

$$\log B_n \sim \log \log n,$$

in place of (8). Hence, the cost of including the regressor $1/t^{1/2}$ grows more slowly than it does when the regressor is stationary. Apparently, the reason for this cost reduction is that as n increases, the model (11) shrinks towards the simpler model $y_t = u_t$, in which there are no coefficients to estimate. Hence, in models like (11), we can get closer to the true model than we could if the regressor were stationary. Note that this is the case even though the rate of convergence of the maximum likelihood estimate of β in (11) is only $\sqrt{\log n}$ rather than $\sqrt{n}$.

Evaporating trends like the regressor $1/t^{1/2}$ in (11) can be useful in modeling intercept creep, where the intercept is allowed to shift from one level to another over time. For instance, in the linear regression model with independent variables x_t

$$y_t = \mu + \frac{\beta}{t^\alpha} + \delta' x_t + u_t, \quad t = 1, ..., n, \tag{12}$$

the intercept shifts from the initial level $(\mu + \beta)$ at $t = 1$ toward a new level (μ) at $t = \infty$, while the coefficients of x_t remain fixed. An empirical example of this type of intercept creep is the NAIRU in the US over the 1990's, which was observed to shift in a downward direction over this period. Such effects seem important in practice, although specifications like (12) have not yet been used in empirical research to capture them.

A more extreme case is provided when $\alpha = 1$ in (11) and the signal from the regressor $1/t$ is even smaller. In this case, $B_n = \sum_{t=1}^{n} \frac{1}{t^2} = O(1)$ and $\log B_n = O(1)$, so that the cost of including the regressor is bounded as $n \to \infty$. Thus, we can get closer to the truth when we estimate model (11) when $a = 1$ than we can when $\alpha = \frac{1}{2}$. Again, the reason is that the true dgp is more closely approximated by the much simpler model $y_t = u_t$ when $\alpha = 1$.

In other ongoing work, the author has been able to show that the same phenomena arises in unidentified structural models, some nearly unidentified models and models with weak instrumentation. In such cases, the bound is again $O_p(1)$, or in cases where there are both identified and nearly unidentified coefficients the inclusion of the nearly unidentified coefficients only introduces an additional cost in the bound that is of $O_p(1)$. Thus, we have the curious outcome that although the coefficients are hard to estimate (Phillips, 1989) and confidence intervals for them may be unbounded (Dufour, 1997), the inclusion of such regressors does not seriously penalize the bound that determines how close we can get to the true dgp or how well we can forecast.

The Bounds are Attainable Asymptotically

The limitation results discussed above provide bounds on how close we can come in empirical modelling to the true dgp and in forecasting to the optimal forecast. It turns out that these bounds are attainable, at least asymptotically. In particular, we can construct empirical models G_n for which

$$\left(-\log \frac{dG_n}{dP_\theta}\right) / \left(\log \det B_n\right) \to_{P_\theta} \frac{1}{2}. \tag{13}$$

One way of attaining the bound asymptotically is to take G_n to be the Bayesian measure $P^n = \int P_\theta \pi(\theta) d\theta$ for any proper Bayesian prior $\pi(\theta)$. We can also use an empirical model Q_n which is based on an asymptotic approximation to P^n and defined by its density

$$\frac{dQ_n}{dP_\theta} = \frac{\pi(\hat{\theta}_n) \exp\left[\ell_n\left(\hat{\theta}_n\right)\right]}{\left(\det B_n\right)^{1/2}}, \tag{14}$$

where $\hat{\theta}_n$ is the maximum likelihood estimate of θ. In the case of improper priors, empirical models G_n may be obtained by taking the conditional Bayes measure, P^{n,n_0}, or its asymptotic approximation Q_{n,n_0}, where the conditioning is on some initial (training) subsample of the data with n_0 observations. Empirical models that are asymptotically equivalent to Q_n and Q_{n,n_0} can also be obtained by prequential methods, like those discussed in Dawid (1984), where we plug in sequentially computed estimates $\hat{\theta}_{t-1}$ of θ into the conditional densities. The reader is referred to Phillips (1996) and Phillips and Ploberger (1996) for details of these constructions and the asymptotic theory associated with them.

When only the model class is known, model selection methods may be used to determine which candidate model is the most appropriate. The density (14) provides a model selection criterion that is consistent (in the sense that the chosen orders converge in probability to the true orders) in a wide range of settings that are useful in econometrics, including unit root testing, determination of the rank of the cointegrating space, lag order determination and trend degree selection. This density is called the PIC density and some of its properties as a model selection device are considered in Phillips (1996), Phillips and Ploberger (1996) and Chao and Phillips (1999). In stationary models, PIC is asymptotically equivalent to the BIC criterion of Schwarz (1978). But in nonstationary models it imposes a higher penalty than BIC and in nearly unidentified models the PIC penalty is weaker. In these respects, PIC has properties that ensure that its use in applications will lead to an empirical model that attains the bounds discussed in the last section, at least asymptotically.

Part 3: One Look to the Future

These properties of PIC model selection and adaptation open up the prospect of using the methods as a basis for automated econometric modeling. In particular, once the model classes are specified, the methods may be employed to find the optimal model amongst the various candidate models in terms of the PIC density (14). The methods were systematically implemented in this fashion by the author to produce automated quarterly forecasts of macroeconomic aggregates 12 quarters ahead for several Asia-Pacific countries over the period 1995-2000, using vector autoregression, reduced rank regression, vector error correction model and Bayesian vector autoregression formats - see Phillips (1995a). The forecasting performance of the methods in this five year experiment turned out to be comparable to that of major macroeconometric forecasting models such as that of Fair's (1994) model of the US economy (see Phillips, 1999). To conduct these exercises the methods were automated in terms of GAUSS programs, following the lines of two earlier applications by the author (1992, 1995b) to historical economic time series for the United States.

One useful feature of the approach is that it offers the flexibility of adaptation of the optimal model on a period by period basis, so that the most suitable model is re-evaluated (including such items as trends and cointegrating rank) as new data becomes available. This approach helps to reduce the impact of misspecification, as discussed earlier, and allows for the model form as well as the estimated coefficients to adapt over time with the arrival of new information. A further advantage is that the order of integration and the cointegrating rank of a system of variables can be monitored and adjusted on a period by period basis, just like other order parameters.

In a recent application of these methods to generate forecasts of New Zealand's real GDP, Schiff and Phillips (2000) showed that this automated approach can produce results that are competitive with the forecasts of professional forecasting institutions. One great advantage of the approach is that these competitive forecasts can be produced almost instantaneously using computer software - all the researcher needs to do is to choose the group of variables to be studied and the model classes to be considered in the application. Schiff and Phillips (1999) also demonstrate how to use these methods to forecast the effects of different economic policies and to evaluate the potential impact of international shocks on domestic economic activity.

Automated methods of this type provide one possible future for the practical use of econometrics. In addition to the author's work described here, their use has been advocated by Hendry (2001) and a software mechanism has been discussed in Hendry and Krolzig (1999, 2001). The approach involves single equation methods that rely on automated significance tests in conjunction with model selection to deal with rival specifications which are unresolved by significance testing. An independent evaluation of the general approach was conducted by Hoover and Perez (1999), leading to broadly favorable conclusions and some recommendations on setting test size more conservatively than the usual 5%. A practical application of the methodology to super 12 rugby attendance modeling is given in Owen and Weatherston (2002).

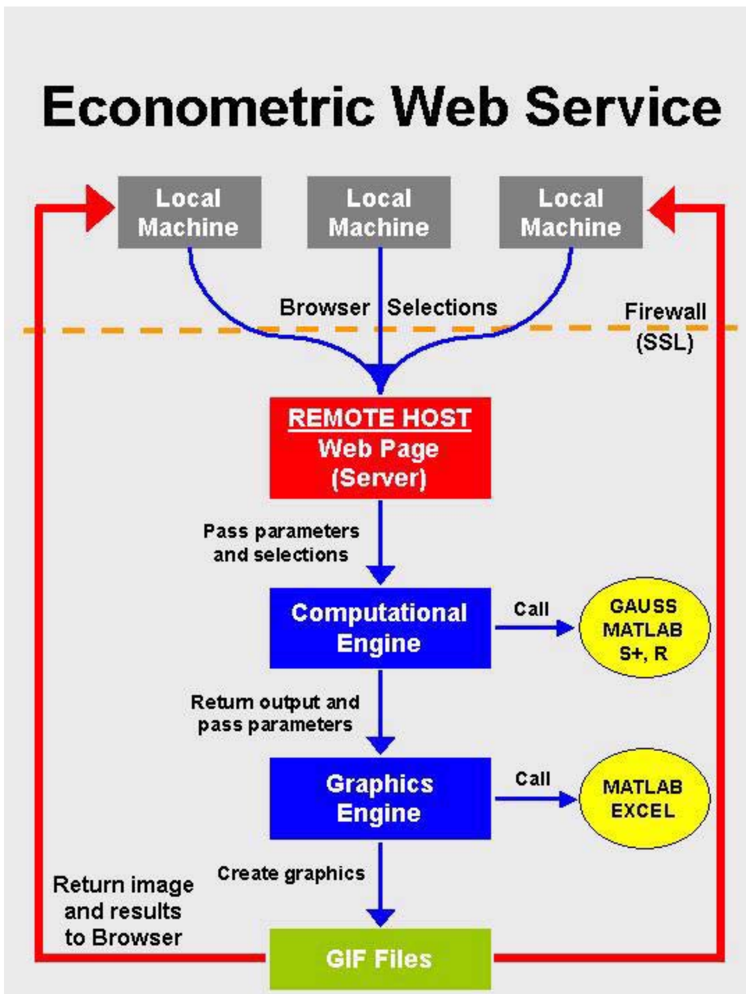

Figure 1: Automated Econometric Computing on the Web

An appealing property of automation is that it can offer econometric modeling methods to a wider community of users. The most direct way in which this service can be accomplished is by means of the internet. Figure 1 outlines a structure that the author has already implemented and tested on a web server that is designed to deliver econometric modeling results and forecasts in response to user activated selections. With this design, local machines can connect to a remote server and by making suitable selections on a web browser a user may estimate models, find the most suitable model in a certain class and use that model for forecasting out to a specified horizon. All of these functions are performed by remote control using programs and data that are resident on the server. For instance, a user may select a variable like GDP, specify the sample period of data to be used, and request forecasts 12 quarters ahead from the most suitable time series model in the class of autoregressions with trend and possible unit roots. The web server responds to this request by passing the selected parameters along to the appropriate statistical software. In the author's implementation, the econometric software is written in GAUSS and the GAUSS engine is used to activate the software from a command embedded in a Visual Basic master program that passes along the user selections. Once GAUSS is activated, the program calls the resident data base for the sample data specified, performs as many regressions as are needed to find the most suitable model in the specified class using as the criterion for model selection the PIC density (14), estimates that model and then uses the model with the fitted coefficients to generate forecasts to the specified horizon. The results are passed back to the master program via parameters that are written into the GAUSS procedure call. Those parameters are used to construct output files and graphics in a suitable format for returning to the user via the web browser. In the author's application, MATLAB is used to construct a graphical display of the sample data and the out of sample projections. These graphics are converted to .gif format, in which form they can be passed along to the user via on the web. Forecasts can also be made under different policy scenarios (for instance, overnight cash rate target settings by the central bank) or different profiles of external shocks (for example, GDP growth rates of a country's major trading partners). The author has been using this process successfully for several years and been able to demonstrate empirical results with a delay of only a few seconds even when the connections to the server are being made over long intercontinental distances.

Perhaps the main advantage of econometric web services of this kind is that they open up good practice econometric technique to a community of users, including unsophisticated users who have little or no knowledge of econometrics and no access to econometric software packages. Much as users can presently connect to financial web sites and see graphics of financial asset prices over user-selected time periods at the click of a mouse button, this software and econometric methodology make it possible for users to perform reasonably advanced econometric calculations in the same way. The web service can be made available on a 24/7 basis so that people can perform the work at their leisure, doing last minute calculations before meetings or even performing online calculations in presentations and lectures. People with no knowledge of econometrics will inevitably have little understanding of the methods

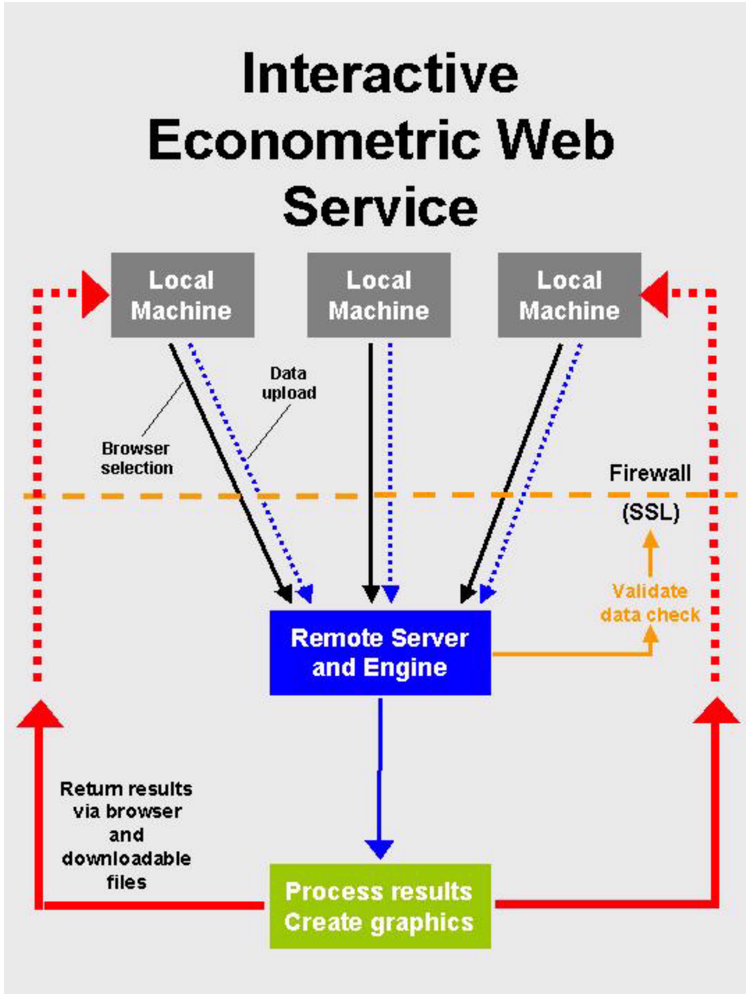

Figure 2: Interactive Econometric Computing on the Web

being used or the limitations of these methods, but confidence bands can be displayed in the forecast profiles and these help to reveal some of the uncertainties of the forecasting and policy analysis process. More sophisticated users can produce forecasts and policy scenarios that can be calibrated against those that have been produced elsewhere. For example, bankers, journalists, business people, politicians and civil servants can obtain forecasts of economic variables relevant to their own work and compare projections under various policy scenarios and external shocks that are of interest to them.

Figure 2 shows how this process can be extended to allow for user supplied data sets. In this case, the user uploads data to the web server and makes selections in the same manner as before. The computation engine on the server then simply uses the supplied data rather than a resident data set in the calculations. One additional difficulty in this interactive form of an econometric web server is that care must be taken to ensure that only data is uploaded through the firewall. This can be accomplished by checking the incoming file to ensure that only numeric data and carriage return characters appear in the file.

Continuing growth in computing power and the extensive use of econometric software packages has made it much easier to do applied econometric work. Web implementations of econometric software of the type just described can be seen as a continuation of this process and they should make econometric methods more generally available and more widely used. While the mechanistic nature of the approach has its limitations, empirical testing of the approach in ex ante forecasting and policy analysis reveals that the approach can work well in practice and can provide competitive forecasts and policy analyses at a very low cost.

Part 4: Afterword

> "*The more we study econometrics, the more there is to discover*"
> Sargan (2003)

It is a truism of any scientific discipline that the more we learn the more there is to know. Like other disciplines, econometrics opens up a maze of complexity as we study it more deeply. The frontier is at once broader in scope and at each point of investigation we continue to discover more fine grain details to resolve. Moreover, as we collect more data and data of different types, we often find that we simply have more to explain and that our understanding of economic behavior does not necessarily improve with larger or even better data sets.

Happily and somewhat characteristically, Denis Sargan himself suggested a partial solution to these problems. The solution takes the form of human capital. Even though automated econometrics of the type described in the last section may play a major role in the practical future of econometrics and even though it is a cliché to say it, new thinkers and new researchers are our greatest ally in moving out the boundaries of econometrics. To wit,

> "*As we discover new problems, we recruit more quality researchers to solve them*" Sargan (2003)

References

Angrist, J. D. and A. B. Krueger (1991). "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106, 979-1014.

Bai, J. and S. Ng (2001). "A PANIC attack on unit roots and cointegration". Working Paper #519, Boston College.

Bai, J. and S. Ng (2002). "Determining the Number of Factors in Approximate Factor Models," forthcoming in *Econometrica.*

Baillie, R.T. (1996), “Long Memory Processes and Fractional Integration in Econometrics”, *Journal of Econometrics* 73, 5-59.

Bandi F. and P. C. B. Phillips (2002). “Fully Nonparametric Estimation of Scalar Diffusion Models,” *Econometrica,* forthcoming.

Baxter, M. and R. G. King (1999). “Measuring business cycles: approximate band-pass filters for economic time series”. *The Review of Economics and Statistics*, 81, 575-593.

Cartwright, N. (1999). *The Dappled World. A Study of the Boundaries of Science.* Cambridge: Cambridge University Press.

Chan, N. H. and C. Z. Wei (1987). “Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics* 15, 1050–1063.

Chao, J. C. and P. C. B. Phillips (1999). “Model selection in partially nonstationary vector autoregressive processes with reduced rank structure.” *Journal of Econometrics*, 91: 227 - 71.

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-Stationary Economic Time Series.* Cambridge, MA: MIT Press.

Clements, M. P. and D. F. Hendry (2001). “Forecasting with difference-stationary and trend-stationary models”. *Econometrics Journal*, 4, S1-S19.

Corbae, D., S. Ouliaris and P. C. B. Phillips (2002). “Band Spectral Regression with Trending Data.” *Econometrica*, 70, 1067-1110.

Dawid, A. P. (1984). “Present position and potential developments: Some personal views, statistical theory, the prequential approach,” *Journal of the Royal Statistical Society*, Series A, 147, 278–292.

Dufour, J-M. (1997). “Some impossibility theorems in econometrics with applications to instrumental variables and dynamic models.” Econometrica, 65, 1365-1388.

Durlauf, S. N. and P. C. P. Phillips (1988). “Trends versus random walks in time series analysis,” *Econometrica* 56, 1333–1354.

Evans, M. K. (2002). *Practical Business Forecasting.* Oxford: Blackwell.

Fair, R. C. (1994). *Testing Macroeconometric Models*”. Cambridge, MA: Harvard University Press.

Fisher, I. (1907). *The Rate of Interest.* New York: Macmillan.

Forchini G. and G. H. Hillier (2002) “Conditional inference for possibly unidentified structural equations”. University of York, mimeographed.

Granger, C. W. J. and P. Newbold (1974). "Spurious regressions in econometrics," *Journal of Econometrics* 74, 111–120.

Hannan, E. J. (1993). "Reducing parameter numbers in econometric modeling." Chapter 16 in P. C. B. Phillips (ed.) *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom.* Oxford: Basil Blackwell.

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators," *Econometrica* 50, 1029–1054.

Härdle, W. and O. Linton (1994). "Applied nonparametric methods." In D. F. McFadden and R. F. Engle III, eds., *The Handbook of Econometrics*, Vol. IV. North–Holland.

Henry, M. and P. Zaffaroni (2001). "The long range dependence paradigm for macroeconomics and finance." forthcoming in *Long range dependence: Theory and applications*, P. Doukhan, G. Oppenheim and M. Taqqu (eds.).

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

Hendry, D. F. (2001). "Achievements and challenges in econometric methodology". *Journal of Econometrics*, 100, 7-10.

Hendry, D.F. and Krolzig, H-M. (1999). "Improving on 'Data Mining Reconsidered' by K.D. Hoover and S.J. Perez." *Econometrics Journal*, 2, 41–58.

Hendry, D.F and H-M. Krolzig (2002). "New Developments in Automatic General-to-specific Modelling." In *Econometrics and the Philosophy of Economics,* edited by B.P. Stigum, MIT Press. Forthcoming

Hoover, K. D. and S. J. Perez (1999). "Data mining reconsidered: encompassing and the general-to-specific approach to specification search". *Econometrics Journal*, 2, 167-191.

Hoover, K. D. (2001). *The Methodology of Empirical Macroeconomics.* Cambridge: Cambridge University Press.

Horowitz, J. L. (1998) *Semiparametric Methods in Econometrics.* New York: Springer.

Jeganathan, P. (1997), "On Asymptotic Inference in Cointegrated Time Series with Fractionally Integrated Errors", *Econometric Theory 15,* 583-621.

Johansen, S. (1988). "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control* 12, 231–254.

Keynes, J. M. (1939). "Professor Tinbergen's method". *Economic Journal*, 49, 558-568.

Kim, J. Y. (1994). "Bayesian asymptotic theory in a time series model with a possible nonstationary process," *Econometric Theory* 10, 764–773.

Kim, C. and P. C. B. Phillips (1999). "Fully modified estimation of fractional cointegration models". Yale University, mimeographed.

Kleibergen, F. (2000). "Pivotal statistics for testing structural parameters in instrumental variables regresssion. " Tinbergen Institute Discussion Paper TI2000-055/4.

Koopmans, T. C. (1937). *Linear Regression Analysis of Economic Time Series*. Haarlem: De Erven F. Bohn.

Koopmans, T. C. and W. Hood (1953). "The estimation of simulataneous linear economic relationships." Chapter 6 in T. C. Koopmans and W. Hood (Eds.) *Studies in Econometric Method*. Cowles Commission Monograph 14, New Haven: Yale University Press.

Lahiri, K. (1999) "ET Interview: G. S. Maddala." *Econometric Theory*, 15, 753-776.

Lancaster, T. (1998). "The Incidental Parameter Problem Since 1948." Mimeographed, Brown University.

Lazarsfeld, P., B. Berelson and H. Gaudet (1944). *The Peoples Choice*. New York: Duell, Sloan and Pearce.

Linton, O. B. (1996) "Second order approximation in the partially linear regression model. " *Econometrica*, 63, 1079-1112.

Maddala, G. S. (1993). *The Econometrics of Panel Data.* Edward Elgar: Aldershot.

Moon, H. and B. Perron (2001). "Testing for a Unit Root in Panels with Dynamic Factors," USC, *mimeo*.

Moon, H. and P. C. B. Phillips (1999). "Maximum Likelihood Estimation in Panels with Incidental Trends." *Oxford Bulletin of Economics and Statistics*, 61, 711-748.

Moon, H. and P. C. B. Phillips (2000)."GMM Estimation of Autoregressive Roots Near Unity with Panel Data" Cowles Foundation Discussion Paper #1274.

Moreira, M. J. (2001). "A conditional likelihood ratio test for structural models". University of California Berkeley, mimeographed.

Nickell, S. (1981): "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417–1426.

Orcutt, G. (1948). "A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States 1919-1932. " *Journal of the Royal Statistical Society*, Series B, 10, 1-45.

Owen P. D. and C. Weatherston (2002) "Uncertainty of outcome and Super 12 attendance: application of a general-to-specific modelling strategy" mimeographed, University of Otago.

Park, J. Y. and P. C. B. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory* 4, 468–497.

Park, J. Y. and P. C. B. Phillips (1989). "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory* 5, 95–131.

Park, J.Y. and P.C.B. Phillips (1999). "Asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, **15**, 269-298.

Park, J.Y. and P.C.B. Phillips (2001). "Nonlinear regressions with integrated time series," *Econometrica*, 69, 117-161..

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics* 33, 311–340.

Phillips, P. C. B. (1987a). "Time series regression with a unit root," *Econometrica,* 55, 277–301.

Phillips, P. C. B. (1987b). "Towards a unified asymptotic theory for autoregression," *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988), "Multiple regression with integrated processes." In N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics* 80, 79–106.

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory* 5, 181–240.

Phillips, P. C. B. (1992). "Bayes Methods for Trending Multiple Time Series with an Empirical Application to the US Economy". Cowles Foundation Discussion Paper #1025.

Phillips, P. C. B. (1995a). "Automated forecasts of Asia-Pacific economic activity". *Asia-Pacific Economic Review*, 1 (1): 92 - 102.

Phillips, P. C. B. (1995b). "Bayesian model selection and prediction with empirical applications". *Journal of Econometrics*, 69: 289 - 331.

Phillips, P. C. B. (1996). "Econometric model determination'. *Econometrica*, 64, 4: 763 - 812.

Phillips, P. C. B. (1998): "Econometric Analysis of Fisher's Equation" Cowles Foundation Discussion Paper, No. 1180. Presented at the Irving Fisher Conference*, Yale University, 1998.*

Phillips, P. C. B. (1998). "New Tools for Understanding Spurious Regressions". *Econometrica*, 66, 1299-1326.

Phillips, P. C. B. (1999). "Forecasts of Asia-Pacific Economic Activity to 2001:4" , *Asia Pacific Economic Review*, Vol. 4, No. 1.

Phillips, P. C. B. (2001). "Descriptive Econometrics for Nonstationary Time Series with Empirical Applications". *Journal of Applied Econometrics*, 16, 389-413.

Phillips, P. C. B. (2002). "New Unit Root Asymptotics in the Presence of Deterministic Trends". *Journal of Econometrics*, forthcoming.

Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. and H.R. Moon (1999). "Linear Regression Limit Theory for Nonstationary Panel Data." *Econometrica,* 67, 1057-1111.

Phillips P. C. B., H. Moon and Z. Xiao (2001). "How to Estimate Autoregressive Roots Near Unity." *Econometric Theory*, 17, 29-69.

Phillips P. C. B. and W. Ploberger (1996). "An Asymptotic Theory of Bayesian Inference for Time Series", *Econometrica,* 64, 381-413.

Phillips P. C. B. and Z. Xiao (1998). "A Primer on Unit Root Testing." *Journal of Economic Surveys*, 12, 423-469.

Phillips P. C. B. and D. Sul (2002). ""Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence" Cowles Foundation Discussion Paper #1362, Yale University.

Ploberger, W. and P. C. B. Phillips (2001). "Rissanen's theorem and econometric time series", in Keuzenkamp, H. A., M. McAleer and A. Zellner (1999). "*Simplicity, Inference and Econometric Modelling*"; Cambridge: Cambridge University Press.

Ploberger, W. and P. C. B. Phillips (2002). "Empirical limits for time series econometric models" *Econometrica* (forthcoming).

Rissanen, J. (1986). "Stochastic complexity and modeling," *Annals of Statistics* 14, 1080–1100.

Rissanen, J. (1987). "Stochastic complexity," *Journal of the Royal Statistical Society*, Series B, 223–239 and 252–265.

Rissanen, J. (1989) *Stochastic Complexity in Statistical Inquiry.* Series in Computer Science, Vol. 15. World Scientific: Singapore.

Robinson, P.M. and D. Marinucci (1998), "Semiparametric Frequency Domain Analysis of Fractional Cointegration", Working Paper 1998/#348, STICERD, London.

Robinson, P.M. and D. Marinucci (1999), "Narrow-Band Analysis of Nonstationary Processes", Working Paper EM/2001/421, London School of Economics.

Sargan, J. D. (1958). "The estimation of economic relationships using instrumental variables," *Econometrica* 26, 393–415.

Sargan, J. D. (1959). "The estimation of relationships with autocorrelated residuals by the use of the instrumental variables," *Journal of the Royal Statistical Society*, Series B, 21, 91–105.

Sargan, J. D. (1983). "Identification and lack of identification". *Econometrica,* 51, 1605-1633.

Sargan, J. D. (2003). "Current problems in econometrics: a personal view." *Econometric Theory*, forthcoming.

Schiff, A. and P. C. B. Phillips (2000). "Forecasting New Zealand's Real GDP." *New Zealand Economic Papers*, 34, 159-182

Schoenberg, , I. J, (1964). "Spline functions and the problem of graduation." Proceedings of the National Academy of Sciences, 52, 333-343.

Schwarz, G. (1978). "Estimating the dimension of a model," *Annals of Statistics* 6, 461–464.

Shimotsu, K. and P. C. B. Phillips (2002). "Exact local Whittle estimation of fractional integration". Cowles Foundation Discussion Paper # 1367.

Sims, C. A. and H. Uhlig (1991). "Understanding unit rooters: A helicopter tour," *Econometrica,*

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Stock, J. H. and M. W. Watson (1998) "Diffusion Indexes." NBER Working Paper #6702.

Stock, J. H. and M. W. Watson (1999) "Forecasting inflation." *Journal of Monetary Economics*, 44, 293-335.

Tinbergen, J. (1939). *Statistical Testing of Business-Cycle Theories*, Vols I and II. Geneva: League of Nations.

Wahba, G. (1978). "Improper priors, spline smoothing, and the problem of guarding against model errors in regression", *Journal of the Royal Statistical Society*, Series B, 40, 364-372.

Wilde, O. (1890). *The Picture of Dorian Gray.* Philadelphia: James Sullivan.

Xiao, Z. and P. C. B. Phillips (1998). "Higher Order Approximations for Frequency Domain Time Series Regression." Journal of Econometrics, 86, 297-336.

Xiao, Z. and P. C. B. Phillips (2002). "Higher Order Approximations for Wald Statistics in Time Series Regressions with Integrated Processes." *Journal of Econometrics*, 2002 (forthcoming).