FACULTY PAPER SERIES

FP 01-10 **October 2001**

Two Approaches to the Model Specification Problem in Econometrics

George C. Davis, Editor
gdavis@tamu.edu

Department of Agricultural Economics
Texas A&M University
College Station, Texas 77843-2124

The Model Specification Problem from a Probabilistic Reduction Perspective

Aris Spanos
Virginia Tech,
Department of Economics,
Blacksburg, VA 24061, USA
TEL: (540)2317981, email: <aris@vt.edu>,

Anya McGuirk
Virginia Tech,
Depts. of Agricultural and Applied Economics & Statistics
Blacksburg, VA 24061, USA
email: <mcguirk@vt.edu>

September 2001

1 Introduction

The primary objective of this paper is to discuss statistical model specification in the context of the Probabilistic Reduction (PR) approach proposed by Spanos (1989,1995a). To understand the problem of specification we need to consider two broad questions:

(a) 'what is being modeled', and

(b) 'what is to be required for an adequate account of model specification'.

From the PR perspective the primary object of modeling is the *actual Data Generating Process* (DGP): the source of the data in coming to inquire about the phenomenon of interest (see Spanos (1986)). The adequacy of any account of model specification will be assessed by its potential in allowing the modeler to learn about the actual DGP and the phenomenon of interest. The manner in which these issues are addressed within the framework of the PR approach sets the stage for understanding and comparing the PR to the traditional textbook and Bayesian approaches.

In section 2 we trace the roots of the statistical model specification problem back to the early 19th century. In section 3 we argue that the PR approach is firmly rooted in the frequentist statistics tradition associated with Fisher and Neyman-Pearson. The PR approach is then viewed as a formalization of Fisher's 'reduction of data' view of

statistics, in section 4. In section 5 we illustrate the strengths of the PR approach using several Monte Carlo simulation experiments.

2 Specification: a brief historical introduction

Statistical model specification, in a primitive form, can be traced back to attempts by Legendre in 1805 and Gauss in 1809 to apply the least-squares method to modeling astronomical and geodesic observations; see Stigler (1986). Gauss re-formulated the least-squares approximation method for 'fitting curves' of the form $y_t = h(\mathbf{x}_t)$, into a *statistical model*, by interpreting the approximation error as a measurement error assumed to be a 'random' variable from a Normal distribution with mean zero and a constant variance. This gave rise to what is known in today's literature as the *Gauss Linear model*:

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \varepsilon_t,\ \varepsilon_t \backsim \mathsf{NIID}(0, \sigma^2),\ t \in \mathbb{T}, \tag{1}$$

where 'NIID' stands for Normal, Independent and Identically Distributed, $\mathbf{x}_t$ denotes a set of *non-stochastic* variables and ε_t denotes 'autonomous' random errors (see Spanos (1986), ch. 18). The conceptual modeling scheme underlying (1) takes the form of the orthogonal decomposition:

$$\boxed{\textbf{observation} \ = \ \textbf{truth} \ + \textbf{error}} \tag{2}$$

where the truth often comes in the form of a 'law'; see Stigler (1999).

For the next century or so the focus of empirical modeling shifted away from relating different variables to studying the distributional structure of individual observable variables of interest using histograms and frequency curves. Viewing this retrospectively, the statistical model of focus was what we nowadays call the *simple Normal model*:

[i] Probability model:
$$\Phi = \left\{ f(x;\boldsymbol{\theta}) = \tfrac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\tfrac{(x-\mu)^2}{2\sigma^2} \right\},\ \boldsymbol{\theta} := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+,\ x \in \mathbb{R} \right\}, \tag{3}$$

[ii] Sampling model: $\mathbf{X} := (X_1, X_2, ..., X_n)$ is a random sample.

It should be noted that the assumptions of Independence (I) and Identically Distributed (ID), defining the notion of a *random sample*, were implicitly imposed because neither concept was clearly understood at that time; these assumptions were formally defined in the 1930s by Kolmogorov and Khintchin. The adoption of the simple Normal model initiated a shift away from attaching the probabilistic structure to errors of measurement ($\varepsilon_t \backsim \mathsf{NIID}(0, \sigma^2)$), to viewing randomness as inherent in the observable variables themselves:

$$X_k \backsim \mathsf{NIID}(\mu, \sigma^2),\ k \in \mathbb{N}.$$

Quetelet popularized this statistical model by applying it to all kinds of sociological data in an attempt to establish the separate discipline of 'social physics' (see Stigler (1986), ch. 5).

In the 1870s Galton, in his attempt to quantify heredity, went beyond the simple Normal model, where observable random variables are modeled separately, to relating two observable random variables (y_t, X_t) via his notion of linear regression:

$$E(y_t \mid X_t = x_t) = \beta_0 + \beta_1 x_t,\ t \in \mathbb{T},$$

where $\beta_0 = E(y_t) - \beta_1 E(X_t)$, $\beta_1 = [Cov(y_t, X_t)/Var(X_t)]$. The *linear regression* became a statistical model based on the conditional distribution $D(y_t \mid \mathbf{X}_t; \boldsymbol{\theta})$:

$$y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + u_t,\ (y_t \mid \mathbf{X}_t = \mathbf{x}_t) \backsim \mathsf{NI}(\boldsymbol{\beta}^\top \mathbf{x}_t, \sigma^2),\ t \in \mathbb{T}, \tag{4}$$

where $u_t = y_t - E(y_t \mid \mathbf{X}_t = \mathbf{x}_t)$. The coefficient β_1 was then used to define the (contemporaneous) *correlation* between two random variables:

$$Corr(y_t, X_t) = \left(\sqrt{\frac{Var(X_t)}{Var(y_t)}}\right)\beta_1 = \frac{Cov(y_t, X_t)}{\sqrt{Var(y_t), Var(X_t)}},\quad t \in \mathbb{T}.$$

providing two very important techniques for relating observable random variables in empirical modeling (see Stigler (1986)).

In the late 19th century Karl Pearson extended the scope of the simple Normal model by introducing numerous new distributions (Gamma, Beta, Pareto, Laplace etc.) via the *Pearson family*. He implicitly retained the IID assumptions but their inappropriateness for some cross-section biological data soon became apparent when crab measurements data gave rise to bimodal histograms, which Pearson explained as due to heterogeneity: two Normal distributions with different means are superimposed; see Stigler (1986). The IID assumptions were also called into question in the 1880s by Lexis under the guise of testing 'the stability' of a time series. The stability of a time series $(y_t,\ t = 1, 2, ..., T)$ was evaluated by comparing it to a *baseline series* of the form:

$$X_k = \theta + \epsilon_k,\ X_k \backsim \mathsf{BinIID}(\theta),\ k \in \mathbb{N},$$

where 'BinIID' stands for Binomial, Independent and Identically Distributed and ϵ_t denotes 'random' errors (see Stigler (1986), ch. 6).

Yule (1897) put forward a direct link between Galton's linear regression model (4) and the Gauss linear model (1) by showing that one can use least-squares to estimate the parameters of both models. Although a major breakthrough in estimation, it unfortunately also contributed significantly to conflating the two models; a confusion that lingers on with terms like 'linear least-squares regression model'.

At the end of the 19th century the lack of understanding of the notion of IID caused many difficulties for modelers of time series data; the IID assumption which is clearly inappropriate for such data led to several well-known problems including *nonsense correlations* (see Yule (1921)). There were numerous ad hoc attempts to handle the temporal dependence and heterogeneity exhibited by such data, but it was not until the late 1920s that Yule and Slutsky put forward more appropriate

statistical models; the *Autoregressive* (AR(p)) and *Moving Average* (MA(q)) models, respectively.

In order to avoid gross anachronisms, it is important to stress that the choice of statistical models considered above belongs to what we nowadays call *descriptive statistics* and not to statistical inference proper. For Karl Pearson statistical model specification amounted to using the first four 'data moments' to choose a frequency curve from the Pearson family that best described the same data.

It is also important to emphasize that before the 1920s the *problem of specification* was practically non-existent because the notion of a statistical model (and its underlying probabilistic assumptions) was not well understood. Arguably, statistical model specification, understood as postulating a statistical model a priori and interpreting the observed data as a realization of the process assumed by that model, was first introduced by Fisher (1922). As the following quotation from "The Design of Experiments" illustrates, Fisher's view of specification was inevitably influenced by his extensive experience with experimental data:

"Statistical procedure and experimental design are only two aspects of the same whole, and that whole comprises all the logical requirements of the complete process of adding to natural knowledge by experimentation" (see Fisher (1935), p. 3)

This enabled Fisher to view the statistical model associated with the experimental design in the context of modeling scheme (2), where $\mathbf{x}_t$ denotes a set of *controlled* (non-stochastic) variables. The modeling strategy being that the onus is on the experimenter to ensure that when the relevant effects, denoted by $\mathbf{x}_t$, *are controlled adequately* the 'remaining' errors, ε_t, are rendered non-systematic.

With Fisher also comes the idea that statistical procedure aims at the actual DGP in a deliberately planned investigation. In this context the 'actual DGP' is inextricably bound up with the experimental design itself and the statistical model is simply a formalization (an error-calibrated form) of the design. As argued in Spanos (1995a), from the statistical model specification viewpoint, Fisher's experimental design and the Gauss linear models are very similar in so far as the actual DGP is assumed to be 'nearly isolated' either by human intervention or 'divine' design. In both situations, scheme (2) provides the conceptual foundation upon which the empirical modeling can be erected.

Unfortunately, the specification of statistical models has progressed very little since the time of Fisher with only occasional discussions in the statistics literature:

"Until relatively recently, the theory has paid little attention to the question of how such a model should be chosen." (see Lehmann (1990), p. 160)

This attitude is aptly caricatured by Dawid (1982) (quoted by Lehmann (1990)):

"Where do probability models come from? To judge by the resounding silence over this question on the part of most statisticians, it seems highly embarrassing. In general, the theoretician is happy to accept that his abstract probability triple $(S, \Im, P(.))$ was found under a gooseberry bush, while the applied statistician's model 'just growed' ."

In view of this, it's not surprising to discover that the current econometric text-

book approach to statistical model specification constitutes an adaptation of the one used in experimental design and least-squares curve fitting, with the stated objective of empirical modeling being 'the quantification of theoretical relationships'. This adaptation, however, is often inappropriate (and misleading) for the statistical analysis of observational (non-experimental) data. This is because the modeling of observational data very rarely 'fits' into the modeling scheme (2) due to the scarcity of observed data that can be realistically viewed as the result of economic 'laws' and/or controlled experiments; see Spanos (1995a).

3 The roots of the PR approach

The primary aim of the Probabilistic Reduction (PR) approach is to put forward a flexible enough framework which (a) can accommodate empirical modeling using both experimental and non-experimental data, (b) enhances the reliability and precision of the inferences drawn, and (c) enables the modeler to learn about the phenomenon of interest. The roots of the PR approach are firmly within the frequency tradition of statistical inference founded by R. A. Fisher (1922,1925,1935) and extended by Neyman and Pearson (1933,1936).

The PR approach reflects the Fisherian insight that 'the process of adding to natural knowledge' turns on the use of data from a deliberately planned analysis in order to learn about the underlying DGP. It adheres, as well, to the fundamental logic by which such learning takes place. A logic in which, according to the frequency tradition, the sampling distribution is a pivotal element in appraising data.

3.1 Fisher's view

Fisher (1922), p. 311, defined the main task of statistics as 'the reduction of a large quantity of data to a few numerical values (parameters); a reduction which adequately summarizes all the relevant information in the original data'. This is repeated in Fisher (1925) where he continues:

"The problems which arise in the reduction of data may thus conveniently be divided into three types:

(i) Problems of **Specification**, which arise in the choice of the mathematical form of the population. This is not arbitrary, but requires an understanding of the way in which the data are supposed to, or did in fact, originate. Its further discussion depends on such fields as the theory of Sample Survey, or that of Experimental Design.

(ii) When the specification has been obtained, problems of **Estimation** arise. These involve the choice among the methods of calculating, from our sample, statistics fit to estimate the unknown parameters of the population.

(iii) Problems of **Distribution** include the mathematical deduction of the exact nature of the distributions in random samples of our estimates of the parameters, and of the other

statistics designed to test the validity of our specification (tests of **Goodness of Fit**)." (see ibid. p. 8)

3.2 Specification

Fisher's view of specification as being confined primarily to data from 'sample surveys' and 'experimental design' is exemplified by his earlier statement:

"As regards problems of specification, these are entirely a matter for the practical statistician, for those cases where the qualitative nature of the hypothetical population is known do not involve any problems of this type. In other cases we may know by experience what forms are likely to be suitable, and the adequacy of our choice may be tested *posteriori*." (see Fisher (1922), p. 16)

His remark that the specification of a statistical model 'requires an understanding of the way the data are supposed to, or in fact, originate' makes an allusion to the 'actual DGP'. The importance of this arises from the fact that the ultimate value of empirical modeling is assessed by whether it contributes to learning about the phenomenon of interest. Indeed, the justification of statistical methods and models is found in their ability to provide systematic strategies for learning from observed data; see Mayo (1996).

There is no denying that bridging the gap between the actual DGP on one side and the theory and statistical models on the other is often a very difficult task. This is because both of these models are idealized descriptions of the actual DGP, qualified by *ceteris paribus* clauses. Fisher's view of specification as being confined primarily to data from 'experimental design' simplifies the task of bridging this gap by utilizing experimental design techniques and controls (randomization, blocking, replication) to operationalize the ceteris paribus clauses and ensure the adequacy of the postulated probabilistic assumptions. The traditional econometric approach to specification amounts to attaching white-noise error terms to theoretical relationships to transform a theory to a statistical model. Not withstanding the appeal of its simplicity, this approach is often inappropriate when modeling observational data. Often the ceteris paribus clauses qualifying the theory model are infeasible, and the gap between the theory concepts and what the data measure cannot be bridged adequately in this simplistic way; see Haavelmo (1944), Spanos (1995a).

To introduce more flexibility into the modeling process, the PR approach (see diagram 1) distinguishes between (a) the (stochastic) phenomenon of interest (the actual DGP), (b) the theory model, and (c) the estimable model. The distinction between (a) and (b) is made to avoid the confusion between the phenomenon itself and our attempts to come to grips with it. The distinction between (b) and (c) is designed to allow for the possibility that certain important aspects of the phenomenon in question might not be observable and thus (b) might not be estimable with the available data. It is important to note that 'data' in this context refers to actual or potential data. It is also interesting to point out that in the case of experimental data, the

estimable model coincides with the experimental model and the hierarchy of models (*primary, experimental* and *data* models) introduced by Mayo (1996) correspond, to a great extent, to the theory, estimable and statistical models above.

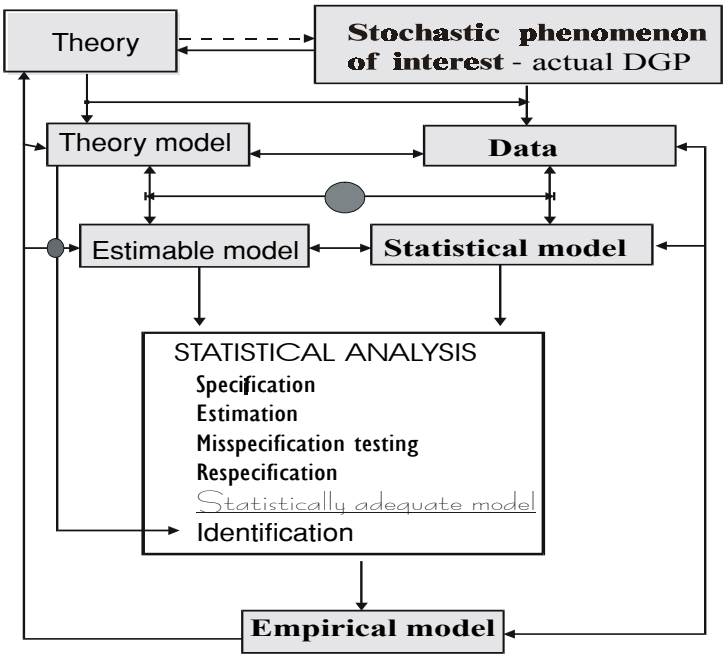

Diagram 1: The Probabilistic Reduction approach

Stochastic phenomenon of interest (Actual DGP): the source of the data in coming to inquire about the phenomenon of interest.

Theory: a conceptual construct purporting to provide an idealized description of the phenomena within its intended scope.

Theoretical model: a mathematical formulation of a theory.

Estimable model: a particular form of the theoretical model which is potentially estimable given the particular observed data chosen.

Statistical model: an internally consistent set of probabilistic assumptions purporting to provide an adequate (probabilistic) 'idealized' description of the stochastic mechanism that gave rise to the observed data with a view to learning about the observable phenomenon of interest.

Empirical model: a reparameterized/restricted form of a statistically adequate statistical model which is interpretable in view of the theory and can be utilized for description, explanation and prediction purposes.

In the context of the PR approach, specification refers to the choice of a statistical model in the context of which the theoretical question of interest will be assessed. The problem facing the modeler is to embed the theory (or substantive question of interest) into a statistical model that adequately accounts for the probabilistic structure of the observed data. This is because the reliability of inference reached on the basis of the estimated statistical model depends crucially on its statistical adequacy; a

misspecified model is likely to give rise to misleading inferences. The embedding the theory within an appropriate statistical model constitutes a most challenging task in empirical modeling because 'appropriateness' is multi-dimensional:

(a) is the statistical model relevant in probing the theory?

(b) are its assumptions satisfied by the data?

(c) are its assumptions internally consistent? and

(d) does it facilitate learning from the data about the phenomenon of interest?

A statistical model is defined in terms of its probabilistic assumptions. For instance, in the case of a simple Normal model the probabilistic assumptions are:

$$\begin{array}{l} \text{[i] } X_k \backsim \mathsf{N}(\mu, \sigma^2),\ (\mu, \sigma^2) \in \Theta := \mathbb{R} \times \mathbb{R}_+,\ k \in \mathbb{N}, \\ \text{[ii] } (X_1, X_2, ..., X_n) \text{ are independent,} \\ \text{[iii] } (X_1, X_2, ..., X_n) \text{ are Identically Distributed} \end{array} \tag{5}$$

The assumptions collectively define what we call a *statistical Generating Mechanism* (GM), which in the case of (5) takes the form:

$$X_k = \mu + u_k,\ k \in \mathbb{N}, \tag{6}$$

where $\mu = E(X_k \mid \mathcal{D}_0)$, $\mathcal{D}_0 = \{S, \varnothing\}$, S being the set of all possible outcomes and $\varnothing$ the empty set. The modeling scheme that can accommodate (6) comes in the form of the orthogonal decomposition:

$$\boxed{X_k = E(X_k \mid \mathcal{D}_k) + u_k, \qquad k \in \mathbb{N}.} \tag{7}$$

The onus being now on the modeler to choose $\mathcal{D}_k$ (the conditioning information set), where $\mathcal{D}_k \subset \mathfrak{F}$, to ensure that the 'remaining' error $u_k = X_k - E(X_k \mid \mathcal{D}_k)$ is non-systematic; the choice of $\mathcal{D}_k$ is influenced by both theory and statistical considerations. This is an extension of (2) to phenomena where there are no 'laws' or controlled experiments; see Spanos (1986, 1995a).

The statistical model is data acceptable when a thorough probing of the probabilistic assumptions reveals no misspecification vis-à-vis the information contained in the data. The assumptions of the model are internally consistent when they are not antithetical. The ultimate success of a statistical model, however, can only be assessed on the basis of whether it facilitates learning from the data about the phenomenon of interest.

3.3 Estimation/Testing within the PR approach

Fisher's second type of problems termed Estimation, should be qualified to "problems of arriving at a suitable statistical procedure" (see Lehmann (1990)), i.e. estimation or testing. Fisher was an avid frequentist and the approach to statistical inference he founded sought estimators and test statistics, which minimize the 'long-run' error probabilities. Taking the Normal model given in (5) with data $\mathbf{x}_n := (x_1, x_2, ..., x_n)$

as an example, Fisher was the first to draw a clear distinction between the unknown *parameters* $\boldsymbol{\theta}:=(\mu,\sigma^2)$ and their *estimators:*

$$\widehat{\mu}(\mathbf{X}_n)=\tfrac{1}{n}\textstyle\sum_{k=1}^{n}X_k,\quad \widehat{\sigma}^2(\mathbf{X}_n)=\tfrac{1}{n}\textstyle\sum_{k=1}^{n}(X_k-\widehat{\mu})^2,$$

The *optimality* of estimators is assessed on the basis of their *inherent reliability* to zero in on the true parameter values $\boldsymbol{\theta}_0:=(\mu_0,\sigma_0^2)$, when the statistical model is true. This reliability is called inherent because it is an attribute that emanates from the structure of the statistical model itself and it is assessed in terms of the sampling distribution of the estimator. The derivation of sampling distributions is classified by Fisher as problems of **Distribution**. In the case of the simple Normal model, the deductive argument is that 'if assumptions [i]-[iii] (see (5)) are true' then the estimators have the following sampling distributions:

$$\widehat{\mu}(\mathbf{X}_n)\backsim \mathsf{N}(\mu,\tfrac{\sigma^2}{n}),\quad \widehat{\sigma}^2(\mathbf{X}_n)\backsim \tfrac{1}{n}\chi^2(n-1).$$

The same deductive argument is involved in deriving a test for the significance of μ based on the test statistic $\tau(\mathbf{X}_n)=\frac{\sqrt{n}\widehat{\mu}(\mathbf{X}_n)}{\sqrt{\widehat{\sigma}^2(\mathbf{X}_n)}}\backsim \mathsf{St}(n-1)$; Student's t with $n-1$ degress of freedom. Its *inherent reliability* is assessed by its ability to detect a false hypothesis; the power of the test. The 'error probabilities' (bias, efficiency, type I and II) involved are usually explained intuitively in terms of the relative frequency of particular errors using the metaphor of a sequence of repetitions of the statistical GM, such as (6), envisaged by the statistical model in question, giving rise to the 'hypothetical' data $\mathbf{x}_n(k)$, $k=1,2,...,N$. For the frequency approach to statistical inference, as perceived in the context of the PR framework, the relevant information function is:

$$g(.;.):\mathcal{X}\times\Theta\to\mathbb{R},$$

which varies over the whole of the sample and parameter spaces. This function gives rise to the *distribution of the sample* $f(\mathbf{x}_n;\boldsymbol{\theta}_0)$ when Θ is confined to $\boldsymbol{\theta}_0$ (the true values) and to the *likelihood function* $L(\mathbf{x}_n^0;\boldsymbol{\theta})$ when $\mathcal{X}$ is confined to the data $\mathbf{x}_n^0$:

$$f(\mathbf{x}_n;\boldsymbol{\theta}_0)=g(\mathbf{x}_n;\boldsymbol{\theta}_0),\ \forall\mathbf{x}_n\in\mathcal{X}\subset\mathbb{R}^n,\qquad L(\mathbf{x}_n^0;\boldsymbol{\theta})=g(\mathbf{x}_n^0;\boldsymbol{\theta}),\ \forall\boldsymbol{\theta}\in\boldsymbol{\Theta}\subset\mathbb{R}^m.$$

Hence, gauging the inherent reliability of the tools we call estimators and tests requires one to consider the probabilities associated with all possible values of the sample, i.e. the whole of the sample space $\mathcal{X}$, as well as the whole of the parameter space. This appeal to the sampling distribution is an essential ingredient of what Fisher, and frequentists in general, regard as necessary for learning about the actual DGP.

By interpreting the observed data as *a realization* of a pre-specified statistical model, Fisher went on to use the latter as a conceptual device in his *significance test reasoning*: the data can be used as evidence against a null hypothesis (specified in terms of the model) insofar as the data are improbably far from what would have been expected if the null were true. Fisher used the same reasoning to test the validity of the assumptions of the statistical model and classified testing 'the validity of our specification' under the same heading of problems of **Distribution**. In modern

terminology Fisher was referring to misspecification testing: the formal assessment of the validity of the probabilistic assumptions specifying the statistical model in question. It should be noted that during the 1920s the 'goodness of fit' test put forward by Karl Pearson was the primary tool for testing distributional assumptions; see Fisher (1925), ch. III. Fisher is explicit about the role of misspecification testing when he argued that: "the adequacy of our choice may be tested posteriori." (see Fisher (1922), p. 16)

In their attempt to improve upon Fisher's significance testing, Neyman and Pearson (N-P) modified his significance testing reasoning (by replacing it with the behavioral decision argument) and narrowed down its intended scope. As argued in Spanos (1999), the N-P testing procedure improved upon Fisher's testing by introducing clear optimality criteria of the same nature as the ones introduced by Fisher in estimation. This, however, was achieved at the price of (inadvertently) limiting testing to within the boundaries of a pre-specified statistical model. Misspecification testing is concerned with establishing the adequacy of the postulated model itself and thus it differs from N-P testing in so far as it probes beyond the boundaries of the pre-specified model:

$$H_0 : f(x) \in \Phi \text{ against } \overline{H}_0 : f(x) \in \mathcal{P} - \Phi,$$

where $\mathcal{P}$ denotes the set of all possible statistical models that can be specified in terms of the joint distribution $D(X_1, X_2, ..., X_n; \boldsymbol{\phi})$. Hence, the latter constitutes testing without the boundaries of the model; see Spanos (1999, 2000) for further discussion concerning the nature of misspecification testing.

In addition to narrowing the scope of Fisher's testing, the N-P procedure replaced his significance testing reasoning with a *behavioral decision rule*. This proved too coarse for empirical modeling purposes and the practice of statistics returned to Fisher's p-value in search for an inferential interpretation of testing; often misinterpreting and abusing the concept. Mayo (1996) put forward a coherent 'inferential' construal of N-P tests by going beyond the inherent reliability of a test and assessing how the observed data $\mathbf{x}_n$ bear upon the particular inference. Her concept of severity provides a post-data evaluation of the inference reached (accept or reject H_0) using counterfactual scenarios.

Interpreting H as denoting either H_0 or H_1, a hypothesis H passes a severe test τ_α with data $\mathbf{x}_n$ if,

(i) $\mathbf{x}_n$ agrees with H, and

(ii) with very high probability, test τ_α would have produced a result that 'fits H less well then $\mathbf{x}_n$ does', if H were false or incorrect.

This reinterpretation of N-P tests deals with the main weakness of the behavioral decision model of N-P tests by extending the pre-data error probabilities, such as size and power, to a 'customized', post-data assessment of the severity with which specific inferences pass the test in question; see Mayo (1996). This sheds additional light on the nature and role of pre-data error probabilities. In scientific contexts the

real value of being able to control error probabilities at small values stems from the fact that it allows the modeler to severely probe, and thereby understand better, the process that gave rise to the observed data $\mathbf{x}_n$; see Mayo and Spanos (2000).

3.4 The Bayesian alternative

The frequentist attitude regarding specification, distribution, and misspecification should be contrasted with the Bayesian approach. For any Bayesian inference concerning $\boldsymbol{\theta}$, invoking the Likelihood principle (see Berger and Wolpert (1984)) implies that the observed data $\mathbf{x}_n$ (via the likelihood function) constitute the only relevant point in the sample space. In fact, Bayesians criticize the use of tail probabilities (p-value, significance level) as their evaluation takes into account other points in the sample space $\mathcal{X}$ beyond the data $\mathbf{x}_n$. Jeffreys (1961) poking fun at the use of p-values stated:

"*What the use of p implies, therefore, is that a hypothesis which may be true may be rejected because it has not predicted observable results which have not occurred.* This seems a remarkable procedure." (see ibid. p.385)

The fact of the matter is that the only way one can assess the reliability of estimators and tests is, indeed, to consider 'observable results which have not occurred'; a modeler should consider their reliability under different circumstances, as described by the distribution of the sample. The Bayesians by confining attention to the likelihood function have no way to assess this inherent reliability. This approach to statistical inference forsakes such assessment in favor of assigning probabilities to the hypotheses (null and alternative) themselves via one's prior distribution $\pi(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. This is then revised in view of the data using the likelihood function via Bayes' theorem to determine the *posterior distribution*:

$$\pi(\boldsymbol{\theta} \mid \mathbf{x}_n^0) \propto \pi(\boldsymbol{\theta}) \cdot L(\mathbf{x}_n^0; \boldsymbol{\theta}),\ \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

The statistical model assumptions enter this revision of subjective beliefs via the likelihood function, and thus, any misspecification is likely to lead to misleading inferences based on $\pi(\boldsymbol{\theta} \mid \mathbf{x}_n^0)$. Can a Bayesian test the statistical adequacy of the model in order to ensure the appropriateness of the likelihood function? The short answer is no, because misspecification testing requires Fisher type significance test reasoning which involves entertaining counterfactual scenarios beyond the observed data $\mathbf{x}_n^0$ and the pre-specified model. Assessing the Bayesian approach on the basis of how it deals with misspecification or/and respecification, we can see that the Bayesian approach leaves no room for revising the statistical model itself; any revisions are within the original model. Hence, unless the Bayesian modeler has the correct specification at the outset, modeling yields no learning about the actual DGP and has no way to get rid of or correct a misspecified model. This raises the important issue of what is being learned from Bayesian modeling: is it mere subjective opinion?

4 PR: formalizing the 'reduction of data'

An important element of the PR approach is the formalization of Fisher's notion of 'the **reduction of data** to a few numerical values; a reduction which adequately summarizes all the *relevant information* in the original data'. The starting point for this formalization is the joint distribution of all the observable random variables involved (the *Haavelmo distribution*); the set of observables has been chosen by some theory in conjunction with what aspects of the phenomenon of interest are measurable. Let all the observables involved be denoted by $\mathbf{Z}_t$ (an $m \times 1$ vector). Kolmogorov's existence theorem (see Billingsley (1986)) ensures that, under certain mild conditions, the probabilistic structure of an observable (vector) stochastic process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$ is fully described by the joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \boldsymbol{\phi})$, for $T > 1$. This distribution demarcates the relevant statistical information because it provides the most general 'description' of the potential information contained in the data. Kolmogorov's theorem warrants the existence of, not only the process itself, but also the 'few numerical values'–the parameters $\boldsymbol{\phi}$. How few these parameters can be depends crucially on the invariance structure of the process. If the actual DGP gives rise to an ever-changing observable processes, its reduction potential is very limited. For the reduction to give rise to applicable models the observable process should enjoy a certain degree of invariance over $t \in \mathbb{T}$. The primary aim of the specification is to capture the invariant features of the phenomenon of interest in the form of the unknown but unchanging parameters $\boldsymbol{\phi}$ of the statistical model. This should be interpreted as requiring that certain measurable aspects of the phenomenon of interest remain invariant or we know how they are changing with t. These regularities are captured by the specified statistical model when its statistical adequacy is established; no departures (misspecifications) from the probabilistic assumptions are detected. The objective in specifying the statistical model is to impose as restrictive (informative) a structure as possible, while ensuring it is in accord with the observed data. The more restrictive the probabilistic structure (when imposed reliably) the higher the precision of the associated inference; see Spanos (2001).

4.1 Reduction assumptions

As shown in Spanos (1999), the probabilistic assumptions that aim at capturing the relevant regularities can be conveniently classified into the three broad categories:
(D) **Distribution,** (M) **Dependence**, (H) **Heterogeneity**.
A statistical model can then be viewed as a reduction from $D(\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_T; \boldsymbol{\phi})$ based on a set of probabilistic assumptions from these three categories.

To make the discussion more concrete, let us consider this reduction in the case of the Normal/Linear Regression (NLR) model as specified in table 1. The vector of observables in this case is: $\mathbf{Z}_t := (y_t, X_t)^\top$, and the reduction assumptions imposed on the process $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$ are: (D) **Normal,** (M) **Independent**, (H) **Identically**

Distributed. The reduction takes the form:

$$D(\mathbf{Z}_1,\mathbf{Z}_2,...,\mathbf{Z}_T;\boldsymbol{\phi}) \overset{\mathsf{I}}{=} \prod_{t=1}^{T} D_t(\mathbf{Z}_t;\boldsymbol{\phi}_t) \overset{\mathsf{IID}}{=} \prod_{t=1}^{T} D(\mathbf{Z}_t;\boldsymbol{\phi}) \overset{\mathsf{IID}}{=} \\ = \prod_{t=1}^{T} D(y_t \mid \mathbf{X}_t;\boldsymbol{\varphi}_1(\boldsymbol{\phi}))\cdot D(\mathbf{X}_t;\boldsymbol{\varphi}_2(\boldsymbol{\phi})),\ \forall(\mathbf{x}_t,y_t)\in\mathbb{R}_X^k\times\mathbb{R}_Y. \tag{8}$$

The details of this reduction are of interest because they bring out the role of each of the reduction assumptions and the reparameterization/restriction from primary parameters $\boldsymbol{\phi}$ to the model parameters $(\boldsymbol{\varphi}_1(\boldsymbol{\phi}),\boldsymbol{\varphi}_2(\boldsymbol{\phi}))$. The imposition of each reduction assumption eliminates successively large subsets of $\mathcal{P}$ (the set of all possible statistical models that can be specified in terms of the joint distribution $D(\mathbf{Z}_1,\mathbf{Z}_2,...,\mathbf{Z}_T;\boldsymbol{\phi})$, eventually reducing it to just one statistical model. In the case of (8), independence reduces the joint to a product of marginal distributions and the ID assumption restricts that to a single distribution with the same parameters. In order to be able to disregard the marginal distribution $D(\mathbf{X}_t;\boldsymbol{\varphi}_2(\boldsymbol{\phi}))$ and concentrate exclusively on $\prod_{t=1}^{T} D(y_t \mid \mathbf{x}_t;\boldsymbol{\varphi}_1)$, we need the reduction assumption of Normality for $\{\mathbf{Z}_t,\ t\in\mathbb{T}\}$.

4.2 Model assumptions

Although the Normal/Linear regression model is specified in terms of assumptions [1]-[8] (see Spanos (1986), it is important to draw the distinction between the *model assumptions* [1]-[8] and the *reduction assumptions*: NIID. As far as statistical inference based on this model is concerned, the relevant assumptions are the model assumptions.

Table 1 − The Normal/Linear Regression model

I. Statistical GM: $y_t=\beta_0+\beta_1x_t+u_t,\ \ t\in\mathbb{T}.$

[1] $\mathcal{D}_t=\{X_t=x_t\}$ is the relevant conditioning information set with
$\mu_t=E(y_t\mid\mathcal{D}_t)=\beta_0+\beta_1x_t$: the systematic component, and
$u_t=y_t-E(y_t\mid\mathcal{D}_t)$: the non-systematic component.

[2] $\boldsymbol{\theta}:=(\beta_0,\beta_1,\sigma^2)$, are the statistical parameters of interest, where
$\beta_0=E(y_t)-\beta_1E(X_t),\quad \beta_1=\frac{Cov(x_t,y_t)}{Var(x_t)},\quad \sigma^2=Var(y_t)-\frac{[Cov(x_t,y_t)]^2}{Var(x_t)}$,
$\boldsymbol{\theta}:=(\beta_0,\beta_1,\sigma^2)\in\Theta:=\mathbb{R}^2\times\mathbb{R}_+$,

[3] X_t is weakly exogenous with respect to $\boldsymbol{\theta}$.

[4] No a priori restrictions on $\boldsymbol{\theta}:=(\beta_0,\beta_1,\sigma^2)$.

[5] $\sum_{t=1}^{T}(x_t-\bar{x})^2\neq 0$; x_t is not constant over $t=1,2,...,T$.

II. Probability model: $\Phi=\{D(y_t\mid x_t;\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta,\ y_t\in\mathbb{R}\}$.

[6]
- (i) $D(y_t\mid x_t;\theta)$ is Normal,
- (ii) $E(y_t\mid X_t=x_t)=\beta_0+\beta_1x_t$ is linear in x_t,
- (iii) $Var(y_t\mid X_t=x_t)=\sigma^2$ is homoskedastic (free of x_t),

[7] The parameters $\boldsymbol{\theta}:=(\beta_0,\beta_1,\sigma^2)$ are t-invariant.

III. Sampling model:

[8] $(y_1,y_2,...,y_T)$ is an independent sample drawn from $D(y_t\mid x_t;\boldsymbol{\theta})$, $t=1,...,T$.

For purposes of specification, misspecification testing and respecification, however, it is of paramount importance to utilize the relationship between the two sets of assumptions as shown in table 2. At the specification stage the relationship between the reduction and model assumptions can be used to assess the adequacy of the statistical model. The model assumptions are more difficult to assess a priori using graphical techniques because they relate to conditional distributions, but the reduction assumptions are very easy to assess because they relate to marginal and joint distributions; see Spanos (1999), ch. 5). Tracing the effects of any departures from the reduction assumptions to the model assumptions will be valuable in choosing the type of misspecification tests to be applied. Respecification can be viewed as tracing the detected departures from the model assumptions to the reduction assumptions and changing the latter accordingly (see Spanos (1986,1999)).

Table 2: **Probabilistic assumptions**

Reduction: $\{\mathbf{Z}_t,\ t\in\mathbb{T}\}$		Model: $\{(y_t \mid \mathbf{X}_t{=}\ \mathbf{x}_t),\ t\in\mathbb{T}\}$
N	$\longrightarrow$	[3], [6](i)-(iii)
I	$\longrightarrow$	[1], [8]
ID	$\longrightarrow$	[2], [7]

5 Specification in empirical modeling

In an attempt to illustrate the PR approach and its potential advantages vis-a-vis the traditional approach, we consider a number of Monte Carlo experiments. These experiments relate to the linear regression model and illustrate the differences between the two approaches at the level of an empirical modeler. All experimental results reported are based on 10,000 replications of sample sizes $T = 50$ and $T = 100$.

For simplicity, let us assume that economic theory suggests that there is a relationship between two variables (X, Y) and the *theory model* takes the initial form:

$$Y = \alpha_0 + \alpha_1 X. \tag{9}$$

In order to draw any inferences regarding this theory, the next step is to embed (9) into a statistical model utilizing data of the form: $\{\mathbf{Z}_t := (x_t, y_t),\ t = 1, 2, ..., T\}$. From the statistical viewpoint, these data are viewed as as realization of a vector stochastic process $\{\mathbf{Z}_t := (x_t, y_t),\ t \in \mathbb{T}\}$, whose probabilistic structure is fully determined by its joint distribution; see Billingsley (1986).

5.1 Experiment 1 - Normal/Linear Regression (NLR)

Consider the case where the observed data $\{(x_t, y_t),\ t = 1, 2, ..., T\}$ are generated via a bivariate IID, Normal processes, with primary parameters:

$$E(Y_t) = 2,\quad E(X_t) = 1,\quad Var(Y_t) = 1,\quad Var(X_t) = 1,\quad Cov(Y_t, X_t) = 0.5 \tag{10}$$

In view of the theory model (9), one could consider the regression model associated with (10) as the embedding statistical model. The information in (10) enables us to deduce that the *true regression model* is:

$$y_t = 1.5 + 0.5x_t + u_t, \quad \sigma^2 = .75, \quad \text{and } \Re^2 = 1 - \tfrac{\sigma^2}{Var(Y_t)} = 0.25, \quad t \in \mathbb{T}. \tag{11}$$

Further, model assumptions [1]-[8] hold by definition and the weak exogeneity of X_t with respect to the model parameters $\boldsymbol{\theta} := (\beta_0, \beta_1, \sigma^2)$, allows us to focus exclusively on the conditional distribution $D(y_t|X_t; \boldsymbol{\theta})$.

We begin by illustrating how PR modelers would proceed on the basis of a particular realization $\{\mathbf{Z}_t := (x_t, y_t),\ t = 1, 2, ..., T\}$. Their first step would be to examine plots of the data in an attempt to *specify* the appropriate statistical model; see Spanos (1999), ch. 5-7. As argued above, one can assess the model assumptions [1]-[8] via the reduction assumptions of NIID for $\{\mathbf{Z}_t,\ t \in \mathbb{T}\}$.

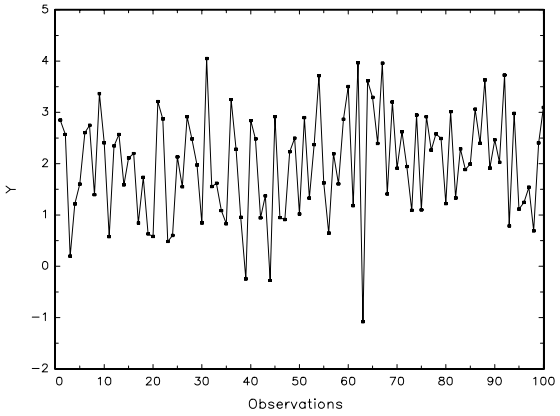

Fig. 1: t-plot of y_t

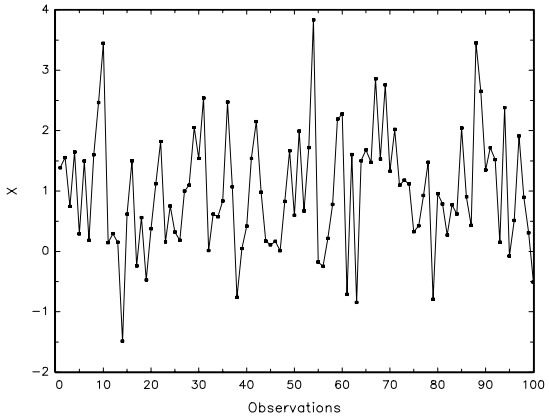

Fig. 2: t-plot of x_t

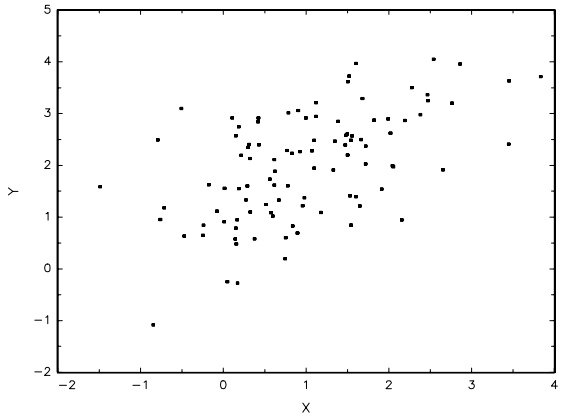

Fig. 3: Scatter-plot of (x_t, y_t)

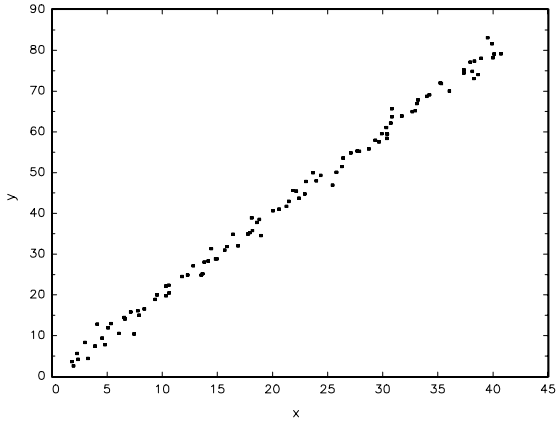

Fig. 4: Scatter-plot of (x_t, y_t)

From the t-plots of y_t and x_t (Figs 1-2) one observes that (a) both the mean and variance of y_t and x_t appear to be constant over t, (b) both processes exhibit independence over t, and (c) the marginal distributions appear to be bell-shape symmetric around a constant mean; a graph of the smoothed histogram of both these variables verifies the bell-shape symmetry; see Spanos (1999), ch. 5. From the scatter-plot of (x_t, y_t) (see fig. 3) one can see the elliptically shaped scatter with a positive principal axis. Thus, a first educated guess would be that the variables are NIID and thus a linear regression model is likely to be appropriate. The estimated regression model using the simulation data is given in Table 3.

The t-statistics reported (τ_{β_i}) are those obtained when testing that the $\beta_i's$ differ from their true values in (11), and they are used as indicators of the reliability of inference. As one can see, the mean estimates are close to their theoretical counterparts and the percentage of rejections (% reject) of the t-test is close to the nominal significance level (α) for β_1. In practice, the PR modeler would not proceed to draw inferences before testing the model assumptions. While there are many possible variants to the misspecification tests that can be applied, the battery of tests implemented by the PR modeler would (a) include tests of all testable assumptions and (b) include individual as well as joint tests of the assumptions. Further, it is important to emphasize that, although misspecification testing is applied in a piece-meal fashion, the null hypothesis being tested is:

$$H_0: \text{ [1]-[8] are valid, vs. } H_1: \text{ any of [1]-[8] are invalid.} \tag{12}$$

Hence, the PR modeler would never interpret the specific test results in isolation, but rather, they are viewed in the context of (12).

Finally, the foundation of misspecification testing within the PR approach is provided by the following lemma.

Conditional expectation orthogonality lemma. Consider the random variable y_t and the random vector $\mathbf{X}_t$ (a $m \times 1$ vector) defined on the same probability space $(S, \mathcal{F}, \mathbb{P}(.))$ such that $E(|y_t|) < \infty$, $E(|\mathbf{X}_t|) < \infty$. Then the only function $g(\mathbf{X}_t)$ for which the following relationship holds for every well behaved (Borel) function $h(\mathbf{X}_t)$:

$$E([y_t - g(\mathbf{X}_t)]h(\mathbf{X}_t)) = 0, \tag{13}$$

is the conditional expectation, i.e. (13) holds iff: $g(\mathbf{X}_t) = E(y_t \mid \sigma(\mathbf{X}_t))$.
NOTE that $\sigma(\mathbf{X}_t)$ denotes the $\sigma-$field generated by $\mathbf{X}_t$. This lemma follows from the orthogonal decomposition (7) and can be easily extended to the higher moments of the conditional distribution $D(y_t \mid \mathbf{X}_t; \boldsymbol{\theta})$:

$$E([u_t^r - E(u_t^r \mid \sigma(\mathbf{X}_t))]h(\mathbf{X}_t)) = 0, \ r = 2, 3, ...$$

These orthogonality conditions can be used to devise misspecification tests for the model assumptions. By choosing $h(\mathbf{X}_t)$ judiciously the modeler can probe for misspecifications by testing for non-orthogonality leading to auxiliary regressions for

misspecification testing. Details of the specific 'auxiliary regression' tests applied in this paper can be found in the Appendix.

Table 3 - True: NLR // Estimated: NLR				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.5015	0.1218	1.5004	0.0869
$\hat{\beta}_1$	0.4985	0.1218	0.4996	0.0869
$\hat{\sigma}^2$	0.7505	0.0207	0.7500	0.0104
R^2	0.2528	0.0898	0.2516	0.0647
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	0.0092	0.0068	0.0036	0.0050
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$	-0.0131	0.0471	-0.0051	0.0485

Table 4 - True: NLR // Estimated: NLR				
	T=50		T=100	
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
D'AP-Normality	1.9980	0.0559	2.0555	0.0601
D'AP-Skewness	-0.0041	0.0965	-0.0040	0.1010
D'AP-Kurtosis	-0.0095	0.1005	-0.0091	0.1058
Durbin-Watson	2.0006	0.1061	2.0007	0.1116
AC Test: $\hat{u}_{t-1}, x_t$	1.0298	0.0468	0.9994	0.0488
AC Test: y_{t-1}, x_{t-1}, x_t	1.0853	0.0534	1.0385	0.0542
White's Homosked.	0.9603	0.0366	0.9618	0.0400
RESET(2) Linearity	1.0803	0.0522	1.0529	0.0533
Joint Mean (A)	1.0718	0.0526	1.0347	0.0521
trend in mean	1.0934	0.0527	1.0467	0.0529
RESET(2) Linearity	1.0900	0.0560	1.0549	0.0570
$\hat{u}_{t-1}$ in mean(1)	1.0719	0.0548	1.0275	0.0518
Joint Mean (B)	1.0670	0.0528	1.0354	0.0509
trend in mean	1.0929	0.0532	1.0457	0.0524
RESET(2) Linearity	1.0912	0.0556	1.0574	0.0559
y_{t-1}, x_{t-1}in mean	1.0611	0.0518	1.0321	0.0514
Joint Variance	0.9449	0.0304	0.9731	0.0425
trend in variance	1.0368	0.0477	1.0265	0.0483
RESET(2) Homosk.	0.9633	0.0404	0.9806	0.0459
ARCH(1)	0.8816	0.0258	0.9340	0.0345

The results from the complete battery of misspecification tests obtained for the

NLR (see table 4), indicate that there are no significant departures from model assumptions [6]-[8] for the data at hand, and thus we can proceed to draw reliable inferences.

In contrast to the PR modeler, a traditional modeler would typically only report the estimated regression results and perhaps a few misspecification tests–usually a Durbin Watson (D-W) test of autocorrelation, and perhaps a test of Normality and/or a test for homoskedasticity (such as White's test). To see how this approach might give rise to unreliable inferences, let us consider generating data which exhibit some departures from the assumptions of experiment 1.

5.2 Experiment 2 - Heterogeneous NLR

Consider the case where the observed data $\{(x_t, y_t),\ t = 1, 2, ..., T\}$ are generated via a bivariate Independent, Normal processes, with heterogeneous means (trending):

$$E(Y_t) = 2 + 0.8t, \quad E(X_t) = 1 + 0.4t, \quad Var(Y_t) = 1, \quad Var(X_t) = 1, \quad Cov(Y_t, X_t) = 0.5 \tag{14}$$

In this case, the true regression model takes the form:

$$y_t = 1.5 + 0.6t + 0.5x_t + u_t, \quad \sigma^2 = .75, \quad \text{and } R^2 = 1 - \frac{\sigma^2}{Var(Y_t)} = 0.25,\ t \in \mathbb{T}. \tag{15}$$

A traditional modeler, in view of (9) is likely to begin by estimating a linear regression model. The results in Table 5 are typical of results one would obtain.

Table 5 - True: NLR with trend // Estimated: NLR, no trend				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	0.4619	0.4501	0.2281	0.3147
$\hat{\beta}_1$	1.9587	0.0401	1.9892	0.0148
$\hat{\sigma}^2$	2.9445	0.3840	2.9853	0.2664
R^2	0.9789	0.0028	0.9945	0.0005
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau^*_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	0.9151	0.1164	0.6454	0.0678
$\tau^*_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	47.554	1.0000	133.82	1.0000
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	-1.9679	0.4757	-3.531	0.9683
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$	35.4064	1.0000	100.18	1.0000
Misspecification Test	Statistic	% reject (.05)	Statistic	% reject (.05)
Normality	2.0238	0.0559	2.0402	0.0568
D-W	1.9984	0.0939	2.0089	0.0260
White's Homosked.	0.9571	0.0396	0.9685	0.0420

Taken at face value, the estimated model looks very reasonable. The R^2 is very high and a simple t-test, $\tau^*_{\beta_1}$, for β_1 indicates that the effect of x_t on y_t is, indeed, significantly different from zero. Further 'confirmation' of how good the results appear satisfactory is given by the scatter plot of the two data series (fig. 4), which seems considerably 'better' than fig. 3 - a clear illusion stemming from misspecification.

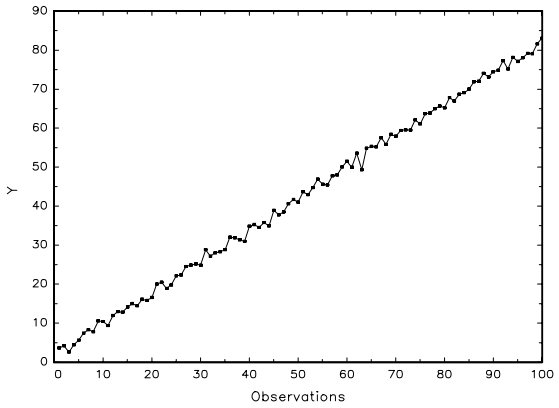

Fig. 5: t-plot of y_t

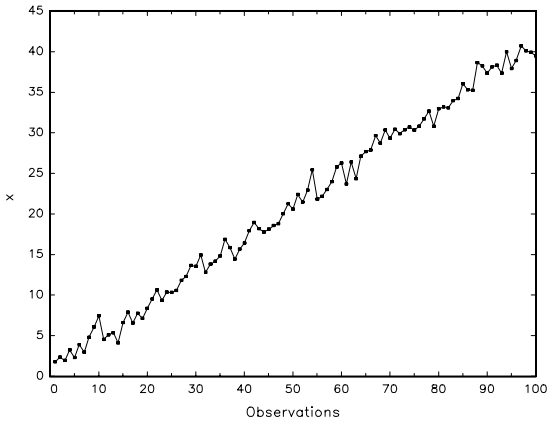

Fig. 6: t-plot of x_t

To understand the extent to which this researcher can be misled, we test whether β_0 and $\beta_1$differ from their *true* underlying values $(\tau_{\beta_i}, i = 0, 1)$. From Table 5, one can see that, for the case of β_1, the true hypothesis is rejected 100% of the time! For the sake of argument, let's suppose that the traditional modeler applies a DW test, a White test of homoskedasticity, and perhaps even a test of normality. The results in Table 5, taken again at face value, do not indicate significant departures from the underlying assumptions; in view of table 5, a traditional modeler would be convinced that the results are reliable despite our knowledge to the contrary. The lesson to be learned is that thorough misspecification testing is imperative if one values the reliability of statistical inference.

In contrast, a PR modeler would first look at t-plots of x_t and y_t (Figs 5-6) where the plots suggests that both data series are trending. Detrending the data and examining the resulting data plots gives results nearly identical to those in Figs. 1-3. Thus, the PR modeler is likely to conclude that a NLR *with a trend* might be the appropriate model. The results in Table 7 and Table 8 indicate clearly that this model is indeed statistically adequate and any inferences based on it will be reliable. The one exception is the estimate of $\mathfrak{R}^2$ which is still misleading as computer programs always take deviations from a constant mean for y_t when estimating $Var(y_t)$ (see McGuirk, et al (1993)). A more appropriate estimator which accounts for the trending mean, $\tilde{R}^2$, is also reported in Table 7 for comparison purposes.

Table 7 - True: NLR with trend // Estimated: NLR with trend				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.5019	0.2516	1.4997	0.1755
$\hat{\beta}_1$	0.4988	0.1232	0.4996	0.0874
$\hat{\gamma}$ (trend)	0.6005	0.0500	0.6052	0.0351
$\hat{\sigma}^2$	0.7505	0.0302	0.7500	0.0150
R^2	0.9947	0.0002	0.9186	0.0000
$\tilde{R}^2$	0.2055	0.0970	0.2285	0.0672
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	0.0078	0.0326	-0.0018	0.0283
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$	-0.0102	0.0462	-0.0051	0.0484
$\tau_\gamma = \frac{\hat{\gamma} - \gamma}{\hat{\sigma}_{\hat{\gamma}}}$	0.0097	0.0476	0.0056	0.0489

Table 8 - True: NLR with trend // Estimated: NLR with trend				
	T=50		T=100	
Misspecification Test	Statistic	% reject (.05)	Statistic	% reject (.05)
D'AP-Normality	2.0046	0.0554	2.0533	0.0582
D'AP-Skewness	0.0025	0.0966	-0.0059	0.1001
D'AP-Kurtosis	-0.0142	0.1021	-0.0119	0.1055
Durbin-Watson	2.0421	0.1181	2.0213	0.1230
AC Test: $\hat{u}_{t-1}, x_t$	1.0709	0.0528	1.0259	0.0521
White's Homosked.	0.9509	0.0368	0.9768	0.0444
RESET(2) Linearity	1.0407	0.0498	1.0478	0.0534
Joint Mean (A)	1.0754	0.0525	1.0410	0.0515
trend^2 in mean	1.0248	0.0485	0.9996	0.0487
RESET(2) Linearity	1.0198	0.0493	1.0015	0.0488
$\hat{u}_{t-1}$ in mean(1)	1.1479	0.0601	1.0667	0.0560
Joint Variance	0.9618	0.0355	0.9795	0.0407
trend in variance	1.0418	0.0447	1.0096	0.0466
RESET(2) Homosk.	1.0441	0.0464	1.0149	0.0467
ARCH(1)	0.9417	0.0301	0.9665	0.0362

What if the PR modeler failed to look at the relevant t-plots and estimated an NLR without taking into account the apparent heterogeneity? The misspecification testing results reported in table 9 would have indicated the problem. These results indicate very clearly (in 100% of the draws!) that heterogeneity in the mean is present. By re-specifying the model to include a trend, and assessing the adequacy of the new

model, the modeler would obtain the results of Tables 7-8 on which valid inferences can be drawn. It is important to emphasize that respecification in the context of the PR approach does not amount to adopting the alternative model in a misspecification test!

Table 9 - True: NLR with trend // Estimated: NLR				
	T=50		T=100	
Misspecification Test	Statistic	% reject (.05)	Statistic	% reject (.05)
D'AP-Normality	2.0238	0.0559	2.0402	0.0568
D'AP-Skewness	-0.0021	0.0986	-0.0005	0.1035
D'AP-Kurtosis	-0.0087	0.1024	-0.0096	0.1042
Durbin-Watson	1.9984	0.0939	2.0089	0.0260
AC Test: $\hat{u}_{t-1}, x_t$	0.9681	0.0413	1.0091	0.0480
AC Test: y_{t-1}, x_{t-1}, x_t	36.1196	1.000	74.0822	1.000
White's Homosked.	0.9571	0.0396	0.9685	0.0420
RESET(2) Linearity	1.0374	0.0475	1.0186	0.0513
Joint Mean (A)	47.0486	1.0000	97.5641	1.000
trend in mean	133.431	1.0000	284.64	1.000
RESET(2) Linearity	1.0453	0.0503	1.0685	0.0560
$\hat{u}_{t-1}$ in mean(1)	1.0529	0.0480	1.0315	0.0535
Joint Mean (B)	35.6778	1.0000	73.4899	1.000
trend in mean	30.4338	1.0000	60.4789	1.000
RESET(2) Linearity	1.0980	0.0547	1.0969	0.0585
y_{t-1}, x_{t-1}in mean	1.0932	0.0536	1.0486	0.0553
Joint Variance	1.1031	0.0571	1.0089	0.0461
trend in variance	1.4582	0.0930	1.1611	0.0646
RESET(2) Homosk.	1.4271	0.0893	1.1472	0.0638
ARCH(1)	0.9101	0.0279	0.9260	0.0342

5.3 Experiment 3 - Non-linear/heteroskedastic Regression

Consider the case where the observed data $\{(x_t, y_t),\ t = 1, 2, ..., T\}$ are generated via a bivariate Gumbel Exponential (GE) with one unknown parameter $\theta = .5$; see Spanos (1999). As indicated above, the PR modeler (with only one eye on the theory model) would use graphical techniques to decide on a reasonable statistical model. The relevant t-plots and scatter plots (Figs 7-9) indicate quite clearly that the data series y_t and x_t are not normally distributed. Both marginal distributions are similar; the distributions are skewed to the right and the t-plots do not exhibit any dependence over t. The apparent asymmetry of the scatter plot suggests that a linear regression function might not be a good model to entertain. A PR modeler is likely to realize that the scatter plot seems to suggest data from a highly skewed distribution, such as

the Log-Normal or Exponential distributions (see Spanos, 1999 for several plots from various distributions). In view of this information, a PR modeler would proceed to estimate the relevant regression and skedastic functions associated with one of these distributions using Maximum Likelihood.

For illustration purposes let us consider the case where the PR modeler, in view of the data plots (fig. 7-9) considers three skewed distributions: the Log-Normal, the Farlie-Gumbel-Morgenstern (F-G-M) Exponential, and the Gumbel-Exponential (GE).

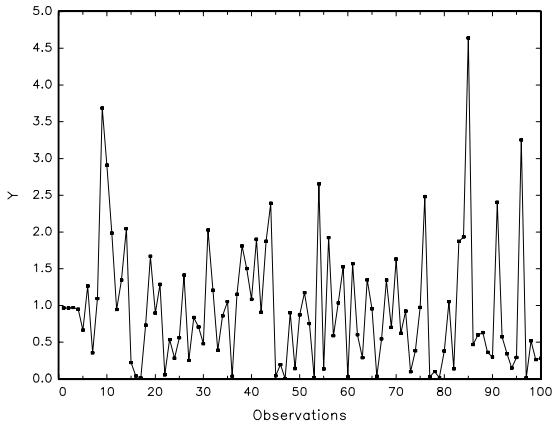

Fig. 7: t-plot of y_t

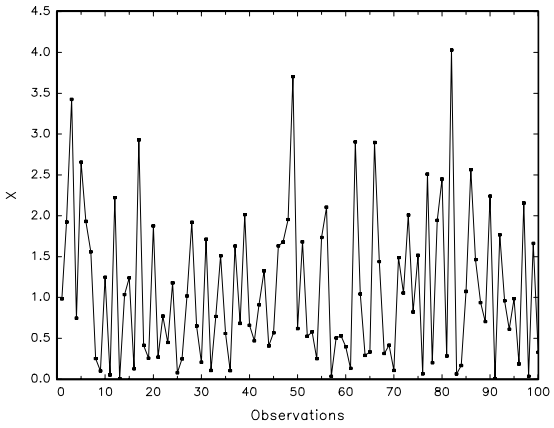

Fig. 8: t-plot of x_t

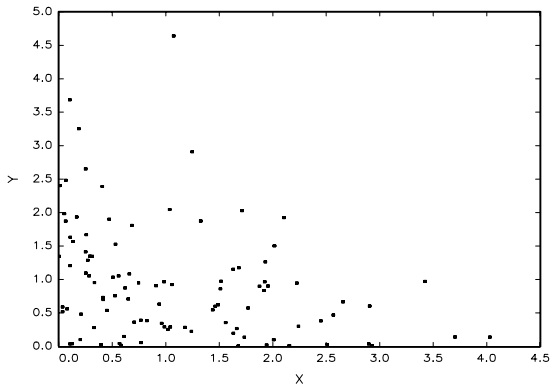

Fig. 9: Scatter-plot of (x_t, y_t)

The Log-Normal regression model. The regression and skedastic functions based on a bivariate Log-Normal distribution take the form (see Spanos (1999)):

$$E(y_t|X_t = x_t) = \left(\frac{x_t}{\mu_2}\right)^{\beta} e^{\mu_1 + \frac{1}{2}\sigma^2}, \qquad Var(y_t|X_t = x_t) = \left(\frac{x_t}{\mu_2}\right)^{2\beta} e^{2\mu_1 + \sigma^2(\exp(\sigma^2)-1)}. \tag{16}$$

Maximum Likelihood Estimation (using simulation data from the GE distribution) of the regression and skedastic functions in (16) yielded the estimates:

$$\underset{(.0268)}{\hat{\mu}_1 = .1379}, \underset{(.0266)}{\hat{\mu}_2 = .1374}, \underset{(.1280)}{\hat{\beta} = -.2963}, \underset{(.3620)}{\hat{\sigma}^2 = .8972}, \quad T = 50;$$
$$\underset{(.0135)}{\hat{\mu}_1 = .0986}, \underset{(.0138)}{\hat{\mu}_2 = .0985}, \underset{(.0847)}{\hat{\beta} = -0.2867}, \underset{(.2526)}{\hat{\sigma}^2 = 0.9081}, \; T = 100.$$

Utilizing the conditional expectation lemma, several new auxiliary regression-based misspecification tests, designed to assess the adequacy of this log-Normal regression model, are proposed; see Appendix. The results of these tests (table 10A) indicate that both the regression and skedastic functions are misspecified.

Table 10A - True: GE regression // Estimated: Log-Normal regression

	T=50		T=100	
Misspecification Test*	Mean	% reject (.05)	Mean	% reject (.05)
Add. non-linearity in mean	-1.2012	0.2662	-1.7363	0.4562
Trend in conditional mean	-0.0046	0.0529	0.0108	0.0467
Mean well-specified: $\alpha_1 = 1$	-6.220	0.5613	-7.268	0.7404
Mean well-specified: $\beta_1 = 1$	-29.202	0.9805	-30.195	0.9991
Add. non-linearity in variance	0.6151	0.0131	0.8772	0.0377
Trend in conditional variance	0.0063	0.0549	-.0128	0.0492
Variance well-specified: $\gamma_0 = 1$	-8.138	0.6814	-9.870	0.8264
Variance well-specified: $\delta_0 = 1$	-4.795	0.6856	-5.832	0.7582

*See Appendix for a more detailed discussion of these tests

The Exponential (Farlie-Gumbel-Morgenstern) (F-G-M) regression model. The regression and skedastic functions based on a bivariate Exponential (F-G-M) distribution take the form (see Kotz, Balakrishnan and Johnson (2000), p. 354):

$$E(y_t|X_t = x_t) = 1 + \tfrac{\alpha}{2} - \alpha e^{-x_t}, \quad Var(y_t|X_t = x_t) = 1 + \tfrac{\alpha}{2} - \tfrac{\alpha^2}{4} - \alpha(1-\alpha)e^{-x_t} - \alpha^2 e^{-2x_t}. \tag{17}$$

Maximum Likelihood Estimation (using simulation data from the GE distribution) of the regression and skedastic functions in (17) yielded:

$$\underset{(.2628)}{\hat{\alpha} = -0.8020}, \; T = 50; \qquad \underset{(.1903)}{\hat{\alpha} = -0.83880}, \; T = 100.$$

This estimated statistical model is also misspecified as the results of table 10B attest; there are clear indications of misspecification for the skedastic function. There are no indications of misspecification for the regression function because for $\alpha = -.99$ (mode of $\widehat{\alpha}$) the regression function in (17) approximates that of the GE (18) very well (see fig. 10). In contrast the skedastic function in (17) does not approximate that in (18); see fig. 11. This brings out the problems of modeling the regression

and skedastic functions *separately* and thus ignoring the fact that they constitute the first two conditional moments of the *same* distribution. The connection between the conditional moments is traditionally ignored when modeling heteroskedasticity; the result is an insufficient utilization of the probabilistic information. It is well known that the first two conditional moments are almost always interrelated when the conditional distribution is derived formally from a proper joint distribution; see Spanos (1999).

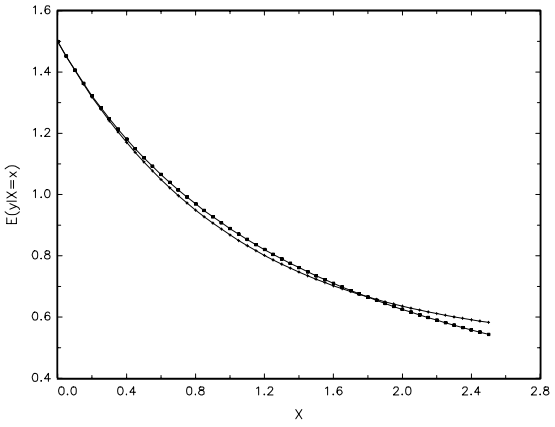

Fig. 10: Regression functions

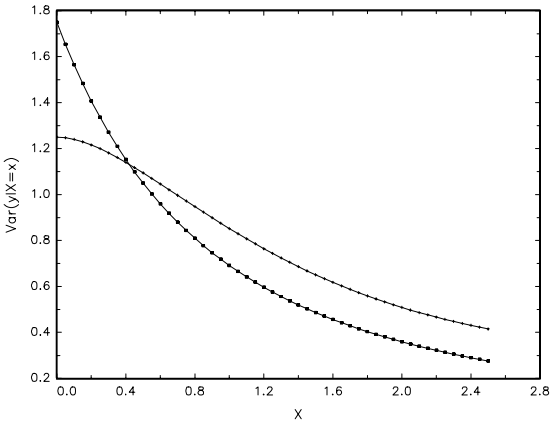

Fig. 11: Skedastic functions

Table 10B - True: GE regression // Estimated: F-G-M regression				
	T=50		T=100	
Misspecification Test*	Mean	% reject (.05)	Mean	% reject (.05)
Add. non-linearity in mean	-.1871	0.0550	-.2447	0.0532
Trend in conditional mean	-0.0009	0.0531	0.0115	0.0466
Mean well-specified: $\alpha_1 = 1$	-.2381	0.0465	-.3086	0.0487
Mean well-specified: $\beta_1 = 1$	-.4586	0.0288	-.5853	0.0330
Add. non-linearity in variance	0.8886	0.1365	1.2033	0.2138
Trend in conditional variance	0.0032	0.0411	.0086	0.0410
Variance well-specified: $\gamma_0 = 1$	-1.3181	0.2494	-1.535	0.3096
Variance well-specified: $\delta_0 = 1$	-.6430	0.1507	-.5441	0.1212
*See Appendix for a more detailed discussion of these tests				

The Gumbel Exponential (GE) regression model. Estimation (using simulation data from the same distribution with $\theta = .5$) of the regression and skedastic functions yield:

$$\begin{array}{cc} \hat{E}(y_t|X_t = x_t) = \frac{1+\widehat{\theta}+\widehat{\theta}x_t}{(1+\widehat{\theta}x_t)^2}, & \hat{V}ar(y_t|X_t = x_t) = \frac{(1+\widehat{\theta}+\widehat{\theta}x_t)^2-2\widehat{\theta}^2}{(1+\widehat{\theta}x_t)^4}, \\ \underset{(.225)}{\widehat{\theta} = .535}, \ T = 50, & \underset{(.158)}{\widehat{\theta} = .515}, \ T = 100. \end{array} \tag{18}$$

When this model is subjected to the new misspecification tests, the results (table 10C) clearly confirm its statistical adequacy. As argued below, the form of non-linearity and heteroskedasticity in this particular case cannot be 'discovered' by ad hoc respecifications. The way the modeler can 'zero in' on such forms of non-linearity and heteroskedasticity is via information relating to the joint distribution underlying $\{(x_t, y_t),\ t = 1, 2, ..., T\}$.

Table 10C - True: GE regression // Estimated: GE regression				
	T=50		T=100	
Misspecification Test*	Mean	% reject (.05)	Mean	% reject (.05)
Add. non-linearity in mean	-0.1066	0.0371	-0.0633	0.0327
Trend in conditional mean	-0.0005	0.0530	0.0121	0.0468
Mean well-specified: $\alpha_1 = 1$	0.0877	0.0242	0.0500	0.0222
Mean well-specified: $\beta_1 = 1$	-0.1026	0.0165	-0.0765	0.0109
Add. non-linearity in variance	0.0694	0.0261	0.0000	0.0176
Trend in conditional variance	0.0013	0.0395	0.0088	0.0417
Variance well-specified: $\gamma_0 = 1$	-0.3549	0.0593	-0.2498	0.0390
Variance well-specified: $\delta_0 = 1$	-0.3235	0.0892	-0.2343	0.0706
*See Appendix for a more detailed discussion of these tests				

Table 11 - True: GE regression // Estimated: NLR				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.2945	0.2229	1.2845	0.1530
$\hat{\beta}_1$	-0.2962	0.1280	-0.2867	0.0847
$\hat{\sigma}^2$	0.9346	0.3770	0.9266	0.2577
R^2	0.0890	0.0500	0.0826	0.0335
t-Statistics	Statistic	% reject (.05)	Statistic	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	6.7548	1.000	9.4787	1.000
$\tau_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	-2.0830	0.5668	-2.9153	0.9166
Misspecification Tests	Statistic	% reject (.05)	Statistic	% reject (.05)
Normality	22.3199	0.9149	42.6088	0.9993
D-W	1.9982	0.1060	2.0007	0.1060
White's Homosked.	1.6306	0.1086	2.4309	0.2708

A traditional modeler commencing with a LRM (in view of the theory model (9)) is likely to have a very hard time diagnosing the particular form of non-linearity and heteroskedasticity associated with the observed data in figures 7-9.

Table 12A - True: GE regression // Estimated: NLR				
	T=50		T=100	
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
D'AP-Normality	22.3199	0.9149	42.6088	0.9993
D'AP-Skewness	3.7087	0.9825	5.2689	1.000
D'AP-Kurtosis	2.3437	0.7012	3.4504	0.9104
Durbin-Watson	1.9982	0.1060	2.0007	0.1060
AC Test: $\hat{u}_{t-1}, x_t$	1.0538	0.0520	1.0203	0.04910
AC Test: y_{t-1}, x_{t-1}, x_t	0.9362	0.0363	0.9617	0.0421
White's Homosked.	1.6306	0.1086	2.4302	0.2708
Linearity: RESET(2)	1.2475	0.0626	1.8822	0.1277
Joint Mean (A)	1.0855	0.0537	1.2850	0.0815
trend in mean	1.0886	0.0513	1.0276	0.0515
RESET(2) Linearity	1.2316	0.0641	1.8456	0.1242
$\hat{u}_{t-1}$ in mean(1)	0.9777	0.0489	0.9851	0.0479
Joint Mean (B)	1.0824	0.0552	1.2233	0.0776
trend in mean	1.0878	0.0521	1.0281	0.0517
RESET(2) Linearity	1.2268	0.0619	1.8374	0.1239
y_{t-1}, x_{t-1}in mean	1.0235	0.0489	1.0112	0.0479
Joint Variance	1.4276	0.0921	1.9827	0.2146
trend in variance	1.0246	0.0376	1.0038	0.0409
RESET(2) Homosk.	2.6111	0.2028	4.2315	0.4645
ARCH(1)	0.6409	0.0221	0.6854	0.0271

Table 12B - True: GE regression // Estimated: NLR				
	T=50		T=100	
Misspecification Test*	Mean	% reject (.05)	Mean	% reject (.05)
Non-linearity in mean	0.7706	0.0662	1.1412	0.1279
Trend in conditional mean	-0.0001	0.0531	0.0117	0.0490
Mean well-specified: $\alpha_1 = 1$	-0.7238	0.0432	-1.0697	0.0892
Mean well-specified: $\beta_1 = 1$	0.0013	0.0000	-0.0020	0.0000
Non-linearity in variance	1.5140	0.2467	1.9825	0.4963
Trend in conditional variance	0.0075	0.0396	0.0024	0.0393
Variance well-specified: $\gamma_0 = 1$	-1.4433	0.1953	-1.8647	0.4211
Variance well-specified: $\delta_0 = 1$	-0.0798	0.0156	-0.0475	0.0143
*See Appendix for a more detailed discussion of these tests				

The results obtained when a LRM is estimated, as shown in table 11, indicate that the Durbin-Watson and homoskedasticity tests are not likely to detect any mis-

specifications, particularly for $T = 50$. The only form of misspecification that will surely show up is that of normality. This is confirmed by the misspecification tests in table 12A, where a more thorough probing indicates clear departures from normality as well some departures from linearity and homoskedasticity; the results of table 12B give a clearer picture of the later departures.

In view of the results in table 11, a traditional modeler who did not do a thorough probing for misspecification, would likely claim that, since normality is the only assumption violated, the inference results are asymptotically valid. Of course, we know in this case, this is not a valid conclusion; not only is normality violated, but so are the assumptions of linearity and homoskedasticity.

Consider now the case where a traditional modeler does do a better job with misspecification testing and notices that the tests in table 12 indicate departures from homoskedasticity. A popular way to 'side-step' the effects of such misspecification is to use heteroskedastically consistent standard errors (HCSE); see White (1980). How would this adjustment affect the reliability of the testing results?

HCSE	T=50		T=100	
t-Statistics	Statistic	% reject (.05)	Statistic	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	6.1496	.9952	8.5633	1.000
$\tau_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	-2.6005	.7929	-3.5838	0.9166

Table 13A - True: GE regression // Estimated: NLR using GLS				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.2136	0.2224	1.2058	0.1618
$\hat{\beta}_1$	-0.2665	0.1449	-0.2544	0.1119
$\hat{\sigma}^2$	1.0315	0.7772	1.0535	0.5916
R^2	0.0661	0.1799	0.0692	0.1730
t-Statistic	Statistic	% reject (.05)	Statistic	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	6.1541	0.9970	8.5224	0.9989
$\tau_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	-2.5514	0.6677	-3.6985	0.8941
Misspecification Test	Statistic	% reject (.05)	Statistic	% reject (.05)
Normality	22.7172	0.9235	43.2381	0.9994
White's Homosk.	1.5496	0.0904	2.3917	0.2471
Linearity-RESET(2)	2.3523	0.1692	3.6775	0.2875

A more 'drastic' way to deal with heteroskedasticity might be to take it into account by using Generalized Least Squares (GLS) where σ_t^2 is estimated using the

fitted values from the auxiliary regression $\widehat{u}_t^2 = \delta_0 + \delta_1 x_t + \delta_2 x_t^2 + v_t$. The results of the GLS estimation (table 13A) suggest that this is clearly not a statistically adequate model. Now, there are not only clear indications of departures from Normality but there are also (increased) indications of non-linearity problems and signs that the heteroskedasticity problem has not been adequately "dealt with"!

In view of the non-linearity, it is interesting to consider a form of respecification often encountered in the context of the traditional approach; the inclusion of x_t^2 in the model. The results from estimating this 2nd degree polynomial model (using GLS) are reported in Table 13B. Interestingly enough, this 'linearity-homoskedasticy corrected' regression model is no more reliable than the original LRM. Moreoever, adding x_t^2 to the original model did not adequately account for the non-linearity; the results indicate lingering non-linearity. This, of course, is not surprising in view of the true regression given in (18). Is this strategy of 'ah hoc correcting' for apparent misspecifications leading the modeler towards the true model? It's very doubtful.

Table 13B - True: GE regression // Estimated: Polyn. regres. (x_t, x_t^2)

	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.3909	0.3302	1.3767	0.2278
$\hat{\beta}_1$	-0.5303	0.5150	-0.5071	0.3276
$\hat{\beta}_2$	0.0697	0.1563	0.0650	0.0830
$\hat{\sigma}^2$	1.1081	0.4583	1.1241	0.3970
R^2	0.1115	0.1722	0.1027	0.1699
t-Statistic	Statistic	% reject (.05)	Statistic	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0}{\hat{\sigma}_{\beta_0}}$	4.8733	0.9887	7.0159	0.9972
$\tau_{\beta_1} = \frac{\hat{\beta}_1}{\hat{\sigma}_{\beta_1}}$	-1.5048	0.3316	-2.3535	0.5710
$\tau_{\beta_2} = \frac{\hat{\beta}_2}{\hat{\sigma}_{\beta_2}}$	0.8624	0.1871	1.3014	0.2767
Misspecification Test	Statistic	% reject (.05)	Statistic	% reject (.05)
Normality	21.1094	0.8925	41.1559	0.9980
Linearity-RESET(2)	2.4566	0.1306	2.3708	0.1493

The estimated regression functions for the GE (highest intercept), OLS-NLR (second highest intercept) and GLS-NLR models are shown in fig. 12, and as we can see, the 'heteroskedasticity-corrected' estimated regression does not improve the approximation! In fig. 13 we can see the estimated regression functions for the the GE (highest intercept) is reasonably well approximated by the GLS-Polynomial (second highest intercept) and F-G-M models, but any inference concerning the regression function remains unreliable.

Consider now the scenario of a PR modeler who failed to examine the relevant

data plots. If the NLR model was estimated in view of (9), thorough misspecification testing (see tables 12A-B) would have revealed departures from normality (both skewness and kurtosis) and some evidence against linearity and homoskedasticity, particularly by the joint tests. In view of this evidence, a PR modeler would return to the data plots for more guidance as to the form of an appropriate regression model.

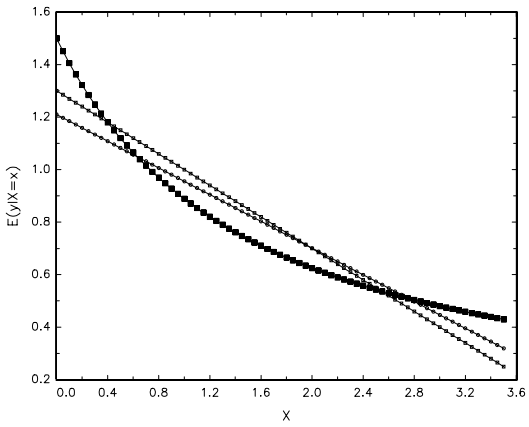

Fig. 12: Estimated Regression functions

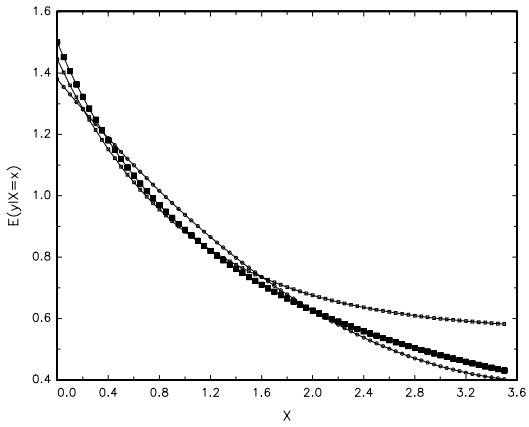

Fig. 13: Estimated Regression functions

5.4 Experiment 4 - Linear Regression with temporal dependence

In this set of experiments we retain the joint normality and stationarity assumptions, but we introduce first-order Markov dependence. For **Experiment 4A** we generate the data with primary parameters:

$$\begin{array}{llll} E(Y_t)=2, & Var(Y_t)=1.250, & Cov(Y_t,Y_{t-1})=0.921, & Cov(Y_t,X_{t-1})=0.754, \\ E(X_t)=1, & Var(X_t)=1, & Cov(X_t,X_{t-1})=0.6, & Cov(Y_t,X_t)=0.846, \end{array}$$

giving rise to the Dynamic Linear regression (DLR(1)) model with:

$$\begin{array}{l} E(y_t|\mathcal{D}_t)=0.7+0.3y_{t-1}+0.5x_t+0.2x_{t-1}, \quad Var(y_t|\mathcal{D}_t)=0.4; \\ \mathcal{D}_t:=\{\sigma(y_{t-1}),X_{t-1}=x_{t-1},\ X_t=x_t\}, \\ x_t=0.4+0.6x_{t-1}+v_t; Var(v_t)=0.64, \qquad y_0 \backsim \mathsf{N}(2,1.250),\ \Re^2=0.68. \end{array} \tag{19}$$

A realization of the data $\{(x_t,y_t),\ t=1,2,...,T\}$ generated by (19) for $T=100$ is shown in figures 14-15.

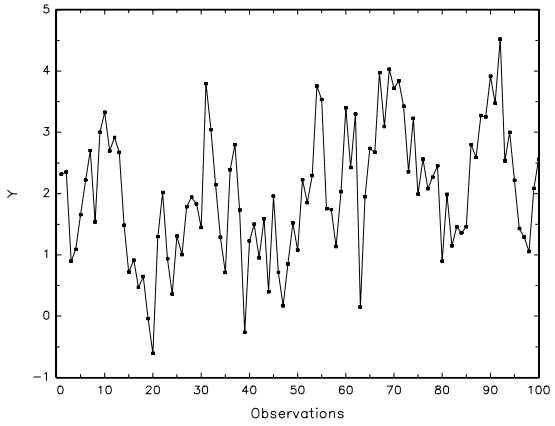

Fig. 14: t-plot of y_t

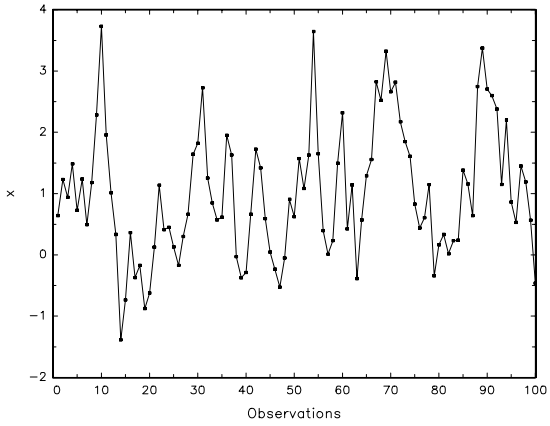

Fig. 15: t-plot of x_t

For **Experiment 4B** we generate the data with primary parameters:

$$E(Y_t) = 2, \quad Var(Y_t) = 1.115, \quad Cov(Y_t, Y_{t-1}) = 0.669, \quad Cov(Y_t, X_{t-1}) = 0.42,$$
$$E(X_t) = 1, \quad Var(X_t) = 1, \quad Cov(X_t, X_{t-1}) = 0.6, \quad Cov(Y_t, X_t) = 0.7,$$

giving rise to the 'restricted' DLR(1) model:

$$\begin{array}{ll} E(y_t|\mathcal{D}_t) = .52 + 0.6y_{t-1} + 0.7x_t - 0.42x_{t-1}, & Var(y_t|\mathcal{D}_t) = 0.4; \\ x_t = 0.4 + 0.6x_{t-1} + v_t; Var(v_t) = 0.64, & y_0 \backsim \mathsf{N}(2, 1.115),\ \Re^2 = 0.64. \end{array} \tag{20}$$

Note, that the true regression model in 4B is considered 'restricted' because it can also be written as:

$$y_t = 1.3 + 0.7x_t + u_t, \quad u_t = 0.6u_{t-1} + v_t,\ t \in \mathbb{T}.$$

That is, for experiment 4B, the *common factor restrictions* implicitly imposed by an error AR(1) process hold.

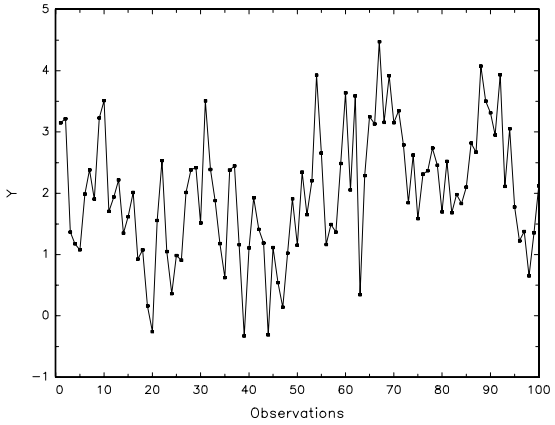

Fig. 16: t-plot of y_t

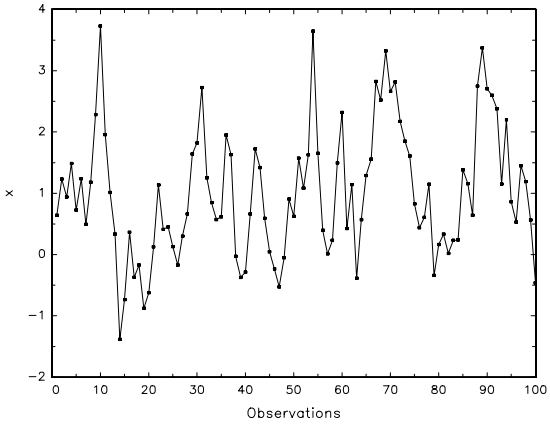

Fig. 17: t-plot of x_t

A realization of the data $\{(x_t, y_t),\ t = 1, 2, ..., T\}$ generated by (20) for $T = 100$ is shown in figures 16-17.

In view of the theory model (9), the traditional modeler would likely begin by estimating theNLR model and perhaps testing the normality, homoskedasticity, and no-autocorrelation assumptions underlying this model; the simulation results for Experiments 4A and 4B are reported in Tables 14A-14B.

Table 14A - True: DLR(1) // Estimated: NLR				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.2816	0.1332	1.2617	0.0920
$\hat{\beta}_2$	0.7185	0.1299	0.7385	0.0910
$\hat{\sigma}^2$	0.5523	0.0848	0.5600	0.0597
R^2	0.4649	0.1161	0.4841	0.0817
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	3.8300	0.9889	5.2535	1.000
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	2.0367	0.5039	3.1534	0.8202
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
Normality	2.0693	0.0593	2.0805	0.0606
Durbin-Watson	1.4518	0.5917	1.4264	0.8731
White's Homosked.	0.9685	0.0370	1.0228	0.0467

Table 14B: True: restricted DLR(1) // Estimated: NLR				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.3104	0.1671	1.3044	0.1169
$\hat{\beta}_2$	0.6896	0.1610	0.6956	0.1149
$\hat{\sigma}^2$	0.5855	0.1255	0.6049	0.0913
R^2	0.4307	0.1297	0.4357	0.0941
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	5.1127	0.9972	7.0978	1.000
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	-0.0950	0.1683	-0.0568	0.1713
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
Normality	2.2376	0.0664	2.4941	0.0861
Durbin-Watson	0.9370	0.9796	0.8674	1.000
White's Homosked.	1.2161	0.0566	1.3099	0.0710

The results in tables 14A-B suggest that temporal dependence would be detected

often enough in both cases. Not surprisingly, despite the similar fit of the two models (based on $\mathfrak{R}^2$), the D-W test is more likely to detect temporal dependence departures in the restricted DLR(1) than in the unrestricted DLR(1).

Note also that, $\hat{\beta}_0$ in experiment 4B is a reasonably good estimator of the intercept in the AR(1) model (with the common factor restrictions imposed) and not the intercept in the (unrestricted) DLR(1) ($\beta_0 = 0.52$), even though the errors have not been modeled as an AR(1) process.

Suppose that, on the basis of the D-W test the traditional modeler (in 4A and 4B) decides to 'correct' the problem by adopting the alternative of an NLR with an AR(1) error model. Tables 15A-15B give simulation results for the models re-estimated using a (2-step) Cochrane-Orcutt correction. As expected, for experiment 4B, where the common factor restrictions hold, inferences regarding β_2 are reliable.

Table 15A - True: DLR(1) // Estimated: restricted DLR(1)

	T=50		T=100	
	Mean	Std	Mean	% reject (.05)
$\hat{\beta}_0$	1.3671	0.1399	1.3601	0.0974
$\hat{\beta}_2$	0.6327	0.1357	0.6399	0.0962
$\hat{\rho}$	0.2589	0.1304	0.2794	0.0898
$\hat{\sigma}^2$	0.4958	0.0624	0.5020	0.0434
R^2	0.3763	0.1228	0.3762	0.0874
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	3.6961	0.9950	5.0930	1.000
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	1.1909	0.2493	1.7246	0.3971

Table 15B - True: restricted DLR(1) // Estimated: restricted DLR(1)

	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	1.30419	0.1176	1.3005	0.0807
$\hat{\beta}_2$	0.6977	0.1156	0.6993	0.0803
$\hat{\rho}$	0.5170	0.1258	0.5592	0.0847
$\hat{\sigma}^2$	0.3925	0.0170	0.3961	0.0081
R^2	0.4483	0.0840	0.4441	0.0575
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	3.5978	0.9822	4.7468	1.000
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	-0.0219	0.0646	-0.0088	0.0563

Further, while a test of whether $\beta_0 = 0.52$ (the true value in the unrestricted

model) gives very misleading results, the test of $\beta_0 = 1.3$ (the true value in the AR(1) formulation) is reliable. However, for experiment 4A – the (unrestricted) DLR(1) model – inferences based on the t-tests of the autocorrelation-corrected model are no more reliable than the ones based on the NLR!

An important question, then is, whether or not the restricted DLR is more likely to be appropriate than the general model, for actual data. If not, the traditional modeler is simply making matters worse by estimating an AR(1) corrected model after a low D-W value. A glance at the implied variance-covariance matrix of $(\mathbf{Z}_t, \mathbf{Z}_{t-1})$, where $\mathbf{Z}_t := (y_t, \mathbf{X}_t)$, when the *common factor restrictions hold* (see Spanos, 1987):

$$Cov(\mathbf{Z}_t, \mathbf{Z}_{t-1}) = \rho Cov(\mathbf{Z}_t),\ t \in \mathbb{T},$$

reveals that it implicitly imposes a constant proportionality between *contemporaneous* ($Cov(\mathbf{Z}_t)$) and *temporal* ($Cov(\mathbf{Z}_t, \mathbf{Z}_{t-1})$) *dependencies* for all the variables involved; highly unrealistic for the overwhelming majority of economic time series data.

It goes without saying that a PR modeler would commence by examining the relevant t-plots (see Fig. 14-15). The positive temporal dependence (in the form of irregular cycles - see Spanos (1999)) is apparent in both plots, but no heterogeneity seems to be present. After 'subtracting' the temporal dependence by estimating AR models for each data series, the t-plots would look very similar to Figs 1-2. Hence, the dynamic linear regression model seems appropriate for these data.

Table 16A - True: DLR(1) // Estimated: DLR(1)

	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	0.7523	0.2006	0.7257	0.1326
$\hat{\beta}_1$	0.2604	0.1232	0.2795	0.0850
$\hat{\beta}_2$	0.4983	0.1140	0.4996	0.0802
$\hat{\beta}_3$	0.2284	0.1432	0.2156	0.1013
$\hat{\sigma}^2$	0.3999	0.0203	0.3999	0.0100
R^2	0.6213	0.0985	0.6356	0.0672
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	0.2005	0.0422	0.1440	0.0320
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$	-0.2884	0.0579	-0.2147	0.0555
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	-0.0143	0.0506	-0.0050	0.0488
$\tau_{\beta_3} = \frac{\hat{\beta}_3 - \beta_3}{\hat{\sigma}_{\beta_3}}$	0.2062	0.0557	0.1596	0.0527

Tables 16A-16B summarize the simulation results obtained by estimating the DLR(1) model. The estimation results seem very accurate and the usual t-tests are likely to be reliable. What is most interesting, however, is that even in cases

where the actual process is 4B, estimating the DLR(1) model yields very reliable inferences. A comparison of the restricted DLR(1) and (unrestricted) DLR(1) results suggests no advantage to estimating the restricted DLR(1) model, even when the restrictions hold! Given the unrealistic nature of the common factor restrictions, and the potential for unreliable inferences when the restrictions do not hold, estimation of error AR(1) type models is not recommended; see also Mizon (1995).

Table 16B - True: restricted DLR(1) // Estimated: DLR(1)				
	T=50		T=100	
	Mean	Std	Mean	Std
$\hat{\beta}_0$	0.6116	0.2048	0.56468	0.1346
$\hat{\beta}_1$	0.5316	0.1248	0.5659	0.0840
$\hat{\beta}_2$	0.6985	0.1143	0.6995	0.0801
$\hat{\beta}_3$	-0.3734	0.1440	-0.3960	0.0994
$\hat{\sigma}^2$	0.3989	0.0207	0.3996	0.0101
R^2	0.6196	0.0920	0.6303	0.0648
t-statistics	Mean	% reject (.05)	Mean	% reject (.05)
$\tau_{\beta_0} = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\beta_0}}$	0.3706	0.0488	0.2609	0.0487
$\tau_{\beta_1} = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}}$	-0.4800	0.0718	-0.3451	0.0625
$\tau_{\beta_2} = \frac{\hat{\beta}_2 - \beta_2}{\hat{\sigma}_{\beta_2}}$	-0.0129	0.0500	-0.0055	0.0490
$\tau_{\beta_3} = \frac{\hat{\beta}_3 - \beta_3}{\hat{\sigma}_{\beta_3}}$	0.3130	0.0677	0.02310	0.0569

For argument's sake, let us suppose that the PR modeler failed to notice the temporal dependence exhibited by the t-plots of $\{(x_t, y_t),\ t = 1, 2, ..., T\}$ and commenced the modeling by estimating a NLR. Testing the model assumptions would reveal the inappropriateness of the NLR model, and the results of misspecification testing would have directed the modeler towards the DLR model; see Tables 17A-17B. It is also interesting to note that these tables suggest that the D-W test, despite its wide use, is not the most probative test to apply, unless the common factor restrictions hold.

In summary, modeling temporal dependence via the error constitutes an inferior way to empirical modeling. Error autocorrelation models (implicitly) reduce temporal dependence among the observable variables to a highly unrealistic form which is rarely true in practice. Even is such rare cases where the common factor restrictions are valid, modeling the temporal dependence directly in terms of the observable random variables constitutes a better strategy as the simulation results in tables 14-17 affirm.

Table 17A - True: DLR(1) // Estimated: NLR				
	T=50		T=100	
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
D'AP-Normality	2.0693	0.0593	2.0805	0.0606
D'AP-Skewness	-0.0035	0.1021	-0.0015	0.1078
D'AP-Kurtosis	-0.0403	0.1075	-0.0239	0.0999
Durbin-Watson	1.4518	0.5917	1.4264	0.8731
AC Test: $\hat{u}_{t-1}, x_t$	4.6189	0.4434	9.4611	0.8209
AC Test: y_{t-1}, x_{t-1}, x_t	10.456	0.9527	20.8686	0.9999
AC Test: $\hat{u}_{t-1}$	1.8923	0.4570	2.9074	0.8215
White's Homosked.	0.9685	0.0370	1.0228	0.0467
RESET(2) Linearity	1.1498	0.0609	1.1689	0.0666
Joint Mean (A)	2.2990	0.2981	3.8719	0.6637
trend in mean	1.2205	0.0701	1.1381	0.0636
RESET(2) Linearity	1.0394	0.0483	1.0124	0.0490
$\hat{u}_{t-1}$ in mean(1)	3.6181	0.3398	8.2911	0.7591
Joint Mean (B)	5.5967	0.8926	10.7654	0.9991
trend in mean	1.1649	0.0599	1.0939	0.0591
RESET(2) Linearity	1.0977	0.0568	1.0387	0.0524
y_{t-1}, x_{t-1}in mean	9.1091	0.9210	19.3985	0.9997
Joint Variance	1.0294	0.0498	1.1965	0.0893
trend in variance	1.0596	0.0482	1.0245	0.0497
RESET(2) Homosk.	0.9761	0.0434	0.9876	0.0471
ARCH(1)	0.9990	0.0428	1.4073	0.0940

Table 17B - True: restricted DLR(1) // Estimated: NLR				
	T=50		T=100	
Misspecification Test	Mean	% reject (.05)	Mean	% reject (.05)
D'AP-Normality	2.2376	0.0664	2.4941	0.0861
D'AP-Skewness	0.0046	0.1206	0.0001	0.1489
D'AP-Kurtosis	-0.1405	0.1128	-0.1294	0.1294
Durbin-Watson	0.9370	0.9796	0.8674	1.000
AC Test: $\hat{u}_{t-1}, x_t$	12.1123	0.9582	48.850	0.9999
AC Test: y_{t-1}, x_{t-1}	21.1752	0.9405	26.0239	1.0000
AC Test: $\hat{u}_{t-1}$	4.3602	0.9607	6.8236	0.9999
White's Homosked.	1.2161	0.0566	1.3099	0.0710
RESET(2) Linearity	1.3763	0.0888	1.4321	0.0961
Joint Mean (A)	8.0123	0.9003	17.0959	0.9996
trend in mean	1.4588	0.0894	1.2406	0.0766
RESET(2) Linearity	1.1198	0.0558	1.0542	0.0550
$\hat{u}_{t-1}$ in mean(1)	16.540	0.9134	43.3482	0.9997
Joint Mean (B)	6.4969	0.8834	13.3428	0.9986
trend in mean	1.3952	0.0833	1.2134	0.0731
RESET(2) Linearity	1.1195	0.0592	1.0518	0.0545
y_{t-1}, x_{t-1}in mean	9.1932	0.8751	22.6862	0.9992
Joint Variance	2.1566	0.2484	4.4256	0.5977
trend in variance	1.0748	0.0528	1.0399	0.0488
RESET(2) Homosk.	1.0482	0.0481	1.0022	0.0460
ARCH(1)	3.3993	0.2664	9.7799	0.6584

6 Conclusion

Focusing primarily on how the questions, 'what is being modeled' and 'what is required for an adequate account of model specification' are addressed, we contrasted the different approaches to empirical modeling. For the Bayesian approach it is not apparent whether the object of modeling is the revision of subjective probabilities or learning about the actual DGP. For the traditional approach the stated object of modeling is 'the quantification of theoretical relationships', mimicking the quantification of 'natural laws'. In terms of learning about the actual DGP via modeling, the Bayesian approach involves no learning unless the true model is known at the outset, because it leaves no room for assessing the adequacy of the statistical model. The traditional approach offers no systematic way to learn about the actual DGP because 'economic laws' are assumed to be known a priori and the modeling focuses primarily on 'confirmation'. In contrast, the PR approach acknowledges the potential

gap between theory models and observed data and provides a systematic framework for learning about the phenomenon of interest by (i) modeling the actual DGP and (i) viewing the statistical model as one of many possible models that can be specified as reductions from the joint distribution of all the observables. The proposed modeling scheme for observational data extends that for experimental data by requiring the modeler to choose the conditioning information set so as to render the remaining errors non-systematic. The Monte Carlo simulation results demonstrate some of the advantages of the PR approach at the level of the practitioner. It is shown that viewing the prespecified statistical model in relation to all possible statistical models, provides a helpful overarching framework that enables the modeler to take informed decisions as to the nature and probabilistic structure of the observed data. Moreover, thorough misspecification testing can detect departures from the assumptions of the specified model as well guide the modeler towards more adequate statistical models.

References

[1] Berger, J. O. and R. L. Wolpert. (1988) *The Likelihood Principle*, 2nd edition. Hayward, CA: Institute of Mathematical Statistics.

[2] Billingsley, P. (1986) *Probability and measure*, 2nd ed. Wiley, New York.

[3] Fisher, R. A. (1922) "On the mathematical foundations of theoretical statistics", *Philosophical Transactions of the Royal Society A*, 222:309-368.

[4] Fisher, R. A. (1925) *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

[5] Fisher, R. A. (1935) *The Design of Experiments*, Oliver and Boyd, Edinburgh.

[6] Haavelmo, T. (1944) "The probability approach to econometrics", *Econometrica*, **12**, suppl., pp. 1-115.

[7] Jeffreys, H. (1961) *Theory of Probability*, 3rd ed., Clarendon Press, Oxford.

[8] Kotz, S., N. Balakrishnan, N. L. Johnson (2000) *Continuous multivariate distributions, vol. 1: models and applications*, 2nd edition, Wiley, New York

[9] Lehmann, E. L. (1990) "Model specification: the views of Fisher and Neyman, and later developments", *Statistical Science*, 5:160-168.

[10] Mayo, D. G. (1996) *Error and the Growth of Experimental Knowledge*, The University of Chicago Press, Chicago.

[11] Mayo, D. G. and A. Spanos. (2000) "A Post-data Interpretation of Neyman-Pearson Methods Based on a Conception of Severe Testing", Measurements in Physics and Economics Discussion Paper Series. LSE.

[12] McGuirk, A., P. Driscoll and J. Alwang. (1993) "Misspecification testing: a comprehensive approach," *Amer. J. Agr. Econ.*, 75:1044-1055.

[13] Mizon G. E. (1995), "A simple message for autocorrelation correctors: Don't", *Journal of Econometrics*, **69**, 1 ,267-288.

[14] Neyman, J. and E. S. Pearson (1933) "On the problem of the most efficient tests of statistical hypotheses", *Phil. Trans. of the Royal Society, A,* 231:289-337.

[15] Spanos, A., (1986) *Statistical foundations of econometric modelling*, Cambridge University Press, Cambridge.

[16] Spanos, A. (1987) "Error autocorrelation revisited: the AR(1) case", *Econometric Reviews,* **6**, pp. 285-94.

[17] Spanos, A. (1989b) "On re-reading Haavelmo: a retrospective view of econometric modeling", *Econometric Theory*, **5**, 405-429.

[18] Spanos, A. (1995a) "On theory testing in Econometrics: modeling with nonexperimental data", *Journal of Econometrics,* 67:189-226.

[19] Spanos, A. (1995b) "On Normality and the Linear Regression model", *Econometric Reviews*, 14:195-203.

[20] Spanos, A. (1999) *Probability Theory and Statistical Inference: econometric modeling with observational data*, Cambridge University Press, Cambridge.

[21] Spanos, A. (2000) "Revisiting data mining: 'hunting' with or without a license," *The Journal of Economic Methodology*, 7:231-264.

[22] Spanos, A. (2001) " Parametric vs. non-parametric inference: statistical adequacy and simplicity," in *Simplicity and Economic Methodology*, eds. A. Zellner, M. McAleer, H. Keuzenkamp, Cambridge University Press.

[23] Stigler, S. M. (1986) *The history of statistics: the measurement of uncertainty before 1900,* Harvard University Press, Cambridge, Massachusetts.

[24] Stigler, S. M. (1999) *Statistics on the Table: the history of Statistical Concepts and Methods*, Harvard University Press, Cambridge, Massachusetts.

[25] White, H. (1980) "A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity", *Econometrica*, **48**, 817–838.

[26] Yule, G.U. (1897) "On the theory of correlation", *Journal of the Royal Statistical Society*, **60**, 812-54. **226**, 267-298.

[27] Yule, G.U. (1921) "On the time-correlation problem", *Journal of the Royal Statistical Society,* **84**, 497-526; reprinted in Stuart and Kendall (1971).

7 Appendix: Misspecification Tests

7.1 NLR Misspecification Tests

For the simulations reported in this paper, we apply several misspecification tests.

Normality: The D'Agostino third sample moment test, the fourth sample moment test proposed by Anscombe and Glynn, and the D'Agostino-Pearson K2 omnibus test combining these two sample moment tests have been shown to have good power over a wide range of alternative distributions (D'Agostino, et al.).

Independence: The four autocorrelation test used include the usual Durbin-Watson, and tests of independence using the following auxiliary regression:

$\hat{u}_t = \beta_0^* + \beta_1^* x_t + \lambda_1 y_{t-1} + \lambda_2 x_{t-1} + v_t.$

$\hat{u}_t = \beta_0^* + \beta_1^* x_t + \lambda_1 \hat{u}_{t-1} + v_t.$

$\hat{u}_t = \beta_0^* + \lambda_1 \hat{u}_{t-1} + v_t.$

Linearity: The functional form test here is a RESET(2) test. We simply use a t-test to assess $\lambda_1 = 0$ in the auxiliary regression: $\hat{u}_t = \beta_0^* + \beta_1^* x_t + \lambda_1 \hat{y}_t^2 + v_t$.

Homskedasticity: White's test of homoskedasticity is used. In the cases, examined here White's test boils down to a basic t-test of the significance of δ in the regression: $\hat{u}_t^2 = \beta_0 + \delta x_t^2 + v_t$.

Parameter Stability: Parameter stability is only assessed in the joint conditional mean and variance tests (see below).

Joint Misspecification Tests

Conditional Mean (A): The joint mean test simultaneously assesses stability of β ($\gamma = 0$), linearity ($\delta = 0$), and independence ($\lambda = 0$) using the following auxiliary regression: $\hat{u}_t = \beta_0^* + \beta_1^* x_t + \gamma t + \delta \hat{y}_t^2 + \lambda \hat{u}_{t-1} + v_t$.

Parameter stability, linearity and independence are then assessed separately by testing the individual components of this joint test one by one. Note that the restricted model in each of these separate tests includes the original x_t as well as the two other test components not being tested.

Conditional Mean (B): The joint mean test simultaneously assesses stability of β ($\gamma = 0$), linearity($\delta = 0$), and independence ($\lambda_1 = \lambda_2 = 0$). It is based on the auxiliary regression: $\hat{u}_t = \beta_0^* + \beta_1^* x_t + \gamma t + \delta \hat{y}_t^2 + \lambda_1 y_{t-1} + \lambda_2 x_{t-1} + v_t$.

Parameter stability, linearity and independence are then assessed separately by testing the individual components of this joint test one by one. Note that the restricted model in each of these separate tests includes the original x_t as well as the two other test components not being tested.

Conditional MeanVariance Test: The joint variance test simultaneously assesses stability of σ^2 ($\gamma = 0$), static homoskedasticity ($\delta = 0$), and dynamic homoskedasticity ($\lambda = 0$; ARCH test) using the following auxiliary regression:

$\hat{u}_t^2 = \beta_0^* + \gamma t + \delta \hat{y}_t^2 + \lambda \hat{u}_{t-1}^2 + v_t.$

Parameter stability, and static and dynamic homoskedasticity are then assessed separately by testing the individual components of this joint test one by one. Note

that the restricted model in each of these separate tests includes the two test components not being tested.

7.2 Exponential Regression Misspecification Tests

Before describing the misspecification tests used to assess the adequacy of the assumptions underlying the Bivariate Exponential model we need to introduce some notation. Define: $\widehat{y}_t \equiv \hat{E}(y_t|X_t = x_t)$, $\widehat{\sigma}_t \equiv \hat{V}ar(y_t|X_t = x_t)$, $\hat{u}_t = y_t - \widehat{y}_t$.

Conditional Mean Tests:

1.) Additional Non-linearities:

The possibility of additional non-linearities in the conditional mean can be assessed by testing $\alpha_2 = 0$ in the following regression: $y_t = \alpha_0 + \alpha_1 \widehat{y}_t + \alpha_2 \widehat{y}_t^2 + u_t$.

2.) Trend in the Conditional mean:

The possibility of a trend in the conditional mean can be assessed by testing $\beta_2 = 0$ in the following regression: $y_t = \beta_0 + \beta_1 \widehat{y}_t + \beta_2 t + u_t$.

3.) Note that not only should we expect $\alpha_2 = 0$ and $\beta_2 = 0$ if the prespecified is the correct model, but we can also expect $\alpha_1 = 1$ and $\beta_1 = 1$. t-tests of these two hypotheses are also reported.

Conditional Variance Tests:

1.) Additional Non-linearities

The possibility of additional non-linearities in the conditional variance can be assessed by testing $\gamma_1 = 0$ in the following regression: $\frac{\hat{u}_t^2}{\widehat{\sigma}_t} = \gamma_0 + \gamma_1 \widehat{y}_t^2 + u_t$.

2.) Trend in the conditional variance:

The possibility of a trend in the conditional variance can be assessed by testing $\delta_1 = 0$ in the following regression: $\frac{\hat{u}_t^2}{\widehat{\sigma}_t} = \delta_0 + \delta_1 t + u_t$.

3.) Note that not only should we expect $\gamma_1 = 0$ and $\delta_1 = 0$ if the Bivariate Exponential model is correct, but we can also expect that $\delta_0 = 1$ and $\gamma_0 = 1$. T-tests of these 2 hypotheses are also reported.