Statistical Adequacy and the Trustworthiness of Empirical Evidence: Statistical vs. Substantive Misspecification

Aris Spanos
Department of Economics,
Virginia Tech, Blacksburg,
VA 24061, USA
e-mail: <aris@vt.edu>

January 2010

Abstract

The paper focuses on how the traditional textbook approach to econometrics, by conflating statistical and substantive information, has contributed significantly to the mountains of untrustworthy evidence accumulated over the last century. In a nutshell, the problem is that when one's favorite theory is foisted on the data, the end result is invariably an empirical model which is both statistically and substantively misspecified, but one has no way to disentangle the two sources of error in order to draw reliable inferences. It is argued that ignoring statistical misspecification, and focusing exclusively on the evaluation of the statistical results — taken at face value — on substantive grounds, has proved a disastrous strategy for learning from data. Moreover, the traditional textbook stratagems of error-fixing designed to alleviate statistical misspecification often make matters worse. Instead, the paper proposes a number of strategies that separate the statistical and substantive sources of information, ab initio, and address the problem by replacing goodness-of-fit with statistical adequacy to secure the statistical reliability of inference, and then proceed to pose questions of substantive adequacy.

1 Introduction

During the 20th century, econometrics has developed from the humble beginnings of using descriptive statistics to 'fit' a line between two variables by least-squares (see Moore, 1914), into an extensive armory of sophisticated statistical tools and procedures for modeling highly complicated dynamic multi-equation systems using a variety of data ranging from time series and cross-section to panel data (see Greene, 2008). Combined with a rapid accumulation of new economic data, together with the widespread use of statistical software on personal computers, the publication of applied econometric papers has been growing exponentially over the last two decades.

Unfortunately for econometrics these impressive technical developments and the accumulation of new and improved data have not been accompanied by any meaningful improvement in the *trustworthiness* of the resulting empirical evidence. If anything, the chronic problem of published untrustworthy empirical evidence seems to have deteriorated as computing power has become more readily available. Indeed, one can make a strong case that as the 21st century unfolds, the applied econometric literature is filled with a compilation of untrustworthy evidence, which collectively provide a completely inadequate empirical foundation for economics; see Spanos (2006a).

This state of affairs in applied econometrics raises several questions including:

[i] What are the main sources of the untrustworthiness of evidence in economics?

[ii] Why does the profession seem rather oblivious to this all-important problem?

[iii] What can practitioners do to alleviate the problem of untrustworthy evidence?

Spanos (2006a) argued that the primary potential sources of error contributing to the untrustworthiness of evidence include:

(I) Inaccurate data. Data $\mathbf{z}_0:=(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n)$ are marred by *systematic errors* imbued by the collection/compilation process.

(II) Incongruous measurement. Data $\mathbf{z}_0$ do not adequately quantify the concepts envisioned by the theory in question; see Spanos (1995).

(III) Statistical inadequacy. The statistical premises of inference are invalid vis-à-vis data $\mathbf{z}_0$.

(IV) Substantive inadequacy. The circumstances envisaged by the theory in question differ 'systematically' from the *actual* data generating mechanism. This inadequacy can easily arise from impractical *ceteris paribus* clauses, external invalidity, missing confounding factors, false causal claims, etc.

The focus of this paper is primarily on statistical vs. substantive inadequacy and how the intermixing of two errors has impeded any real progress in learning from data and gave rise to mountains of untrustworthy evidence. Summed up, the problem is that when one's favorite theory is foisted on the data, the end result is invariably an empirical model which is both statistically and substantively misspecified, but one has no way to disentangle the two sources of error. It is argued that the traditional textbook stratagems designed to sidestep statistical misspecification often make matters worse. The quintessential example is the so called *omitted variables* problem in econometrics, where the search for substantively significant explanatory variables will be led astray by unreliable inference procedures due to statistical misspecification; see Spanos (2006b). This paper proposes a number of strategies to disentangle the two issues by securing statistical adequacy using thorough Mis-Specification (M-S) testing before posing questions of substantive adequacy.

Section 2 introduces the problem of statistical model validation and discusses the possible reasons for its neglect in econometrics. Section 3 provides a brief discussion of certain key foundational problems associated with frequentist model-based inference in an attempt to place the problem of statistical model validation in the broader context of frequentist inference. Section 4 discusses the question of delineating the statistical from the substantive premises of interest using the Linear Regression model as the primary example. Section 5 considers a simple empirical example in an attempt to shed light on some of the important issues raised in sections 2-4.

2 Model validation and empirical evidence

2.1 The Pre-Eminence of Theory (PET) perspective

The single most important contributor to the untrustworthiness of empirical evidence in economics is the methodological framework that has dominated empirical modeling in economics since Ricardo (1817). This framework, known as the Pre-Eminence of Theory (PET) perspective, asserts that empirical modeling takes the form of constructing simple idealized models which capture certain key aspects of the phenomenon of interest, with a view to use such models to shed light or even explain economic phenomena, as well as gain insight concerning potential alternative policies. From the PET perspective empirical modeling is strictly theory-driven with the data playing only a subordinate role in quantifying theory-models (assumed to be true). Empirical modeling in economics invariably begins with a theory model which is then transformed into an econometric model by selecting the'best' available data corresponding to the theoretical variables of interest and attaching white-noise error terms — carrying the essential probabilistic structure needed for inference purposes.

The all-familiar result of foisting one's favorite theory on the data is usually an estimated model which is both *statistically* and *substantively misspecified*, but one has no way to distinguish between the two sources of misspecification and apportion blame. The statistical misspecification [invalid statistical premises vis-a-vis the data] undermines the *reliability* of any inductive inference by inducing a significant discrepancy between the relevant *nominal* and *actual* error probabilities. The surest way to draw an invalid inference is to apply a .05 significance level test when its actual – due to misspecification – type I error probability is closer to .99; see Spanos and McGuirk (2001). This implies that the estimated model cannot be used as a basis of reliable inferences to assess the validity of the substantive information of interest. That is, any inferential claim concerning the sign, magnitude and significance of estimated coefficients, however informal, is likely to be misleading because of the discrepancy between the actual and nominal error probabilities. Hence, when such inferences are drawn they shed no real light on the underlying economic phenomena and no learning from data takes place.

This systemic error-unreliability of inference, however, has not stopped an ever increasing stream of empirical papers filling up an accumulating number of applied econometrics journals claiming that data provide evidence for their pet theories in one way or another. It is no exaggeration to describe the untrustworthiness of accumulating empirical results in economics as the quiet scandal of econometrics that has persistently undermined the empirical foundations of economics for more than a century with no end in sight. This is primarily, but not exclusively, because very few (less than 1%) empirical studies are based on validated (vis-à-vis $\mathbf{z}_0$) statistical premises (Spanos, 2006a).

2.2 Claiming knowledge stemming from strong beliefs?

This unacceptable state of affairs has been wholly exposed by the recent discussions pertaining to the financial crises that burst onto the scene in September 2008. The economists participating in the debate concerning what policies are likely to be most effective in dealing with the deepening recession were invariably invoking *causal knowledge* between key policy variables, like government expenditure, and macro aggregates like the Gross Domestic Product (GDP). The problem was that all they had to offer as *evidence for* their claimed knowledge was a strong *belief* in the appropriateness of their particular economic perspective (classical, Keynesian, monetarist, etc.), and some armchair empiricism based on analogical reasoning from broad tendencies in these variables during past 'similar' episodes.

What does it take to establish reliable causal knowledge? This short answer is a lot more than strong beliefs and anecdotal evidence. In particular, one needs to guard against a variety of potential errors including securing the statistical and substantive adequacy of the models one invokes. To the best of my knowledge, no such empirical models exist in the econometric literature. Very few, if any, such models have been statistically validated and even fewer have passed thorough probing of potential errors to secure *substantive adequacy*: the estimated model provides a *veritable explanation* for the phenomenon of interest.

2.3 Why has model validation been neglected in economics?

In light of the serious implications of statistical misspecification, why has the validation of the statistical premises been so gravely neglected by the applied econometrics literature? There are several reasons which can be ultimately traced to the inadequacy of the PET perspective and the methodological framework it gives rise to.

The *first* reason for this neglect is that practicing econometricians are likely to have a hard time distinguishing between the statistical and substantive premises underlying their inductive inferences. This is because the theory-driven modeling, advocated by the PET perspective, often renders the statistical premises implicit and almost never specified in the form of assumptions that are testable vis-à-vis the data. For instance, in the case of a simultaneous equations model, the statistical premises are defined in terms of the implicit reduced form. But no textbook in econometrics or applied published paper seems to be aware of the simple fact that when the estimated reduced form is statistically inadequate, all inference based on the estimated structural form are questionable at best! Instead, the statistical premises are usually specified indirectly in terms of an incomplete set of assumptions concerning error terms that represent an amalgam of statistical and substantive presuppositions, rendering the separation of the two premises impossible. What makes matters worse is that advocates of the PET perspective often invoke the inevitable substantive unrealisticness of their structural models (Mäki, 2009) as an alibi for ignoring the statistical misspecification problem, when in fact the two problems are fundamentally different (Spanos, 2009a). This will provide the focus of this paper.

Related to the unrealisticness issue is the *second* reason for the neglect of statistical model validation. The erroneous, but widely held, impression that a certain degree of statistical misspecification is both endemic and unavoidable, and the only way to protect the reliability of one's inferences is to weaken the statistical premises and rely on asymptotic inference; see Matyas (1999). Moreover, traditional econometricians often invoke certain generic robustness to 'small' departures from the model assumptions and believe that a large enough sample (n) will smooth out any unreliability of inference. As demonstrated in Spanos (2009b), both the invoked robustness and the large sample remedy rely primarily on wishful thinking because certain 'small' departures sometimes lead to a sizeable discrepancy between nominal and actual error probabilities and certain departures exacerbate the unreliability of inference as n increases. Econometric journals are increasingly being filled with highly technical papers proving *consistency* and *asymptotic Normality* for "yet another estimator" under revised but seemingly less restrictive technical conditions which are often nontestable vis-a-vis the data. Indeed, this explains the fascination of econometric textbooks with Gauss-Markov type results that avoid distributional assumptions, and the increasing allurement of nonparametric techniques. Take for instance the commonly used replacement of a distributional assumption, say Normality, with the nonparametric kernel smoothing assumption that the *unknown* density function $f(x)$ has continuous derivatives of up to order 3; see Li and Racine (2006). This is often presented misleadingly as a choice between making a distributional assumption or not, when in fact the trade off is between an assumption you can easily test with a restriction you have no way of assessing. Normality is easily testable vis-a-vis the data, but how is one supposed to test these restrictions on an unknown $f(x)$? The inference results based on such nontestable assumptions can be highly misleading because they provide ostensibly respectable ways to hide the untrustworthiness of evidence problem behind technical gimcrackery. Although this particular issue is beyond the scope of this paper, Spanos (2001, 2006a, 2010a) makes a case on how misleading such a strategy can be when the trustworthiness of evidence is a primary objective.

The *third* reason for disregarding statistical model validation is the prevailing erroneous perception in econometrics that Mis-Specification (M-S) testing, as a way to secure statistical adequacy, is undefendable against several methodological charges including: (i) illegitimate data-mining, (ii) spurious data-snooping, (iii) pre-test bias, (iv) illicit double-use of data, (v) infinite regress, (vi) circularity and (vii) erroneous diagnoses; see Kennedy (2008). The conventional wisdom in textbook econometrics is that, since the specification of the estimable model is theory-driven, one is supposed to intentionally ignore the chance regularities in the data, since any form of data-snooping could (somehow) bias the inferential claims one can draw from the data. These methodological issues are addressed in Spanos (2000, 2010b).

3 Model-based inference and its reliability

The statistical inference procedures currently employed in econometrics give the impression that their foundations are primarily based on the Fisher-Neyman-Pearson (F-N-P) approach to frequentist inference. A closer examination of the textbook econometric methodological framework, however, reveals that, despite the wholesale adoption of Fisher's estimation and the Neyman-Pearson's testing methods, some of the key tenets of the F-N-P perspective, such as its emphasis on statistical regularities, have been largely ignored in empirical modeling in economics. One of the primary obstacles has been the problem of striking the correct balance between the substantive subject matter and statistical information and their respective roles in empirical modeling; see Spanos (2007b, 2008).

3.1 Frequentist inference and its foundational problems

Fisher (1922) initiated an incisive change of paradigms in statistics by recasting the then dominating *Bayesian-oriented induction by enumeration* (Pearson, 1920), relying on large sample approximations, into a *frequentist model-based induction*, relying on *finite sampling distributions*. He proposed to view the data $\mathbf{z}_0:=(z_1, z_2, ..., z_n)$ as a realization of: (a) a 'random sample' from (b) a pre-specified 'hypothetical infinite population' and made the initial choice (specification) of the statistical model a response to the question:

"Of what population is this a random sample?" (Fisher, 1922, p. 313), emphasizing that: 'the adequacy of our choice may be tested posteriori' (ibid., p. 314).

Fisher's notion of a prespecified statistical model can be formalized as a parametrization of the stochastic process $\{Z_k,\ k\in\mathbb{N}\}$ underlying data $\mathbf{z}_0$. Using the joint *distribution of the sample* $\mathbf{Z}:=(Z_1, ..., Z_n)$, $f(\mathbf{z};\boldsymbol{\theta})$, to provide a general description of the probabilistic structure of $\{Z_k,\ k\in\mathbb{N}\}$, one can define a *statistical model* by:

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})=\{f(\mathbf{z};\boldsymbol{\theta}),\ \boldsymbol{\theta}\in\Theta\},\ \mathbf{z}\in\mathbb{R}^n_Z,\ \text{for}\ \Theta\subset\mathbb{R}^m,\ m<n. \tag{1}$$

$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is chosen to provide an idealized description of the mechanism that generated data $\mathbf{z}_0$ with a view to appraise and address the substantive questions of interest. The basic idea is to construct statistical models using probabilistic assumptions that 'account for' the chance regularities in the data aspiring to adequately account for the underlying data-generating mechanism; see Spanos (1999).

The quintessential *example* of a statistical model is *the simple Normal model*:

$$\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z}){:}\ Z_k \backsim \mathsf{NIID}(\mu,\sigma^2),\ \ \boldsymbol{\theta}{:=}(\mu,\sigma^2){\in}\mathbb{R}{\times}\mathbb{R}_+,\quad k{=}1,2,...,n,..., \tag{2}$$

where '$\backsim$ $\mathsf{NIID}(\mu,\sigma^2)$' stands for 'distributed as Normal, Independent and Identically Distributed, with mean μ and variance σ^2. Note that all the statistical information in $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is encapsulated by: $f(\mathbf{z};\boldsymbol{\theta}){=}(\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp\{-\frac{1}{\sigma^2}\sum_{k=1}^{n}(z_k-\mu)^2$.

Fisher (1925, 1934) established the optimal *estimation* theory of model-based frequentist inference almost single-handedly and Neyman and Pearson (N-P) (1933) supplemented Fisher's significance testing with an optimal *hypothesis testing* theory. Although the formal apparatus of the Fisher-Neyman–Pearson (F–N–P) statistical inference was largely in place by the late 1930s, the nature of the underlying *inductive reasoning* was beclouded in disagreements. Fisher argued for 'inductive inference' spearheaded by his significance testing (Fisher, 1955), and Neyman argued for 'inductive behavior' based on N–P testing (Neyman, 1956). All the same, neither account gave an adequate answer to the basic question:

when do data $\mathbf{z}_0$ provide evidence for (or against) a hypothesis or a claim H?

Indeed, several crucial **foundational problems** were left largely unanswered:
[i] the role of pre-data vs. post-data error probabilities (Hacking, 1965),
[ii] *the fallacy of acceptance* [*no* evidence against the null is misinterpreted as evidence *for* it], and *the fallacy of rejection* [evidence *against* the null is misinterpreted as evidence *for* a specific alternative] (Mayo, 1996),
[iii] when does a model 'account for the regularities' in the data? (Spanos, 1986),
[iv] what is the role of substantive information in statistical modeling?

These foundational problems created endless confusions in the minds of practitioners, especially in the social and medical sciences, concerning the appropriate use and interpretation of the frequentist approach to inference. In economics, there is the additional problem that compounds the undermining of the estimated models:
[v] the huge gap between economic theory concepts and the corresponding available data; see Spanos (1995).

The current methodological framework in econometrics, dominated by the PET perspective, ignores problems [i]-[ii] and addresses [iii] using primarily goodness-of-fit measures. [iv] is dealt with by giving theory veto power over all modeling decisions. In an effort to address these foundational problems more adequately, Spanos (2006a) called for a new *methodological framework* which:
(a) brings out the *gap between theory and data,*
(b) encourages *probing for potential errors* at all stages of modeling,
(c) specifies statistical models in terms of the observable processes,
(d) separates the statistical and substantive information, and
(e) affords the data 'a life of their own' by emphasizing statistical adequacy in place of goodness-of-fit.

At the heart of this new framework, known as *error statistics* (Mayo, 1996, Mayo and Spanos, 2010b), is a refinement/extension of the F-N-P frequentist approach. The *modus operandi* of error statistics is *ascertainable* error probabilities. *Pre-data* error probabilities are used to assess both the optimality and reliability of inference methods. *Post-data* error probabilities, based on *severity* reasoning, provide an *evidential* interpretation for frequentist testing that can be used to address the problems [i]-[ii]; Mayo and Spanos (2006). A crucial pre-condition for rendering these error probabilities ascertainable in practice is the validation of $\mathcal{M}_{\theta}(\mathbf{z})$.This paper focuses primarily on how the error statistical framework can be used to address problems [iii]-[iv].

3.2 The pivotal role of a statistical model

The notion of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ plays a pivotal role in statistics because:

(i) it specifies the inductive premises of inference,

(ii) it determines what constitutes *a legitimate* event,

(iii) it assigns probabilities to all legitimate events via $f(\mathbf{z};\boldsymbol{\theta}),\ \mathbf{z}\in\mathbb{R}^n_Z$,

(iv) it defines what are legitimate hypotheses and/or inferential claims,

(v) it determines the relevant error probabilities in terms of which the optimality and reliability of inference methods is assessed, and

(vi) it designates what constitute legitimate data $\mathbf{z}_0$ for inference purposes.

Formally an event is legitimate when it belongs to the σ-field generated by $\mathbf{Z}$ (Billingsley, 1995). Legitimate data come in the form of data $\mathbf{z}_0$ that can be realistically viewed as a truly typical realization of the process $\{Z_k,\ k\in\mathbb{N}\}$, as specified by $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$. Legitimate hypotheses and inferential claims are invariably about the data-generating mechanism (DGM) and framed in terms of the unknown parameters $\boldsymbol{\theta}$. Moreover, the optimality (effectiveness) of the various inference procedures is ascertained in terms of the relevant error probabilities which are governed by $f(\mathbf{z};\boldsymbol{\theta})$.

One of the most crucial weaknesses of model-based inference since the 1920s has been articulated succinctly by C. R. Rao (2004):

> "The current statistical methodology is mostly model-based, without any specific rules for model selection or validating a specified model." (p. 2)

The methodological framework proposed in Spanos (1986, 1989, 1995, 2010a-b) aims to remedy this weakness as it relates to the foundational problems [iii]-[iv]. The primary objective is to develop a methodology of *Specification, Mis-Specification (M-S) testing* and *Respecification* with a view to secure *statistical adequacy.* That involves how to specify and validate statistical models, how to probe model assumptions, isolate the sources of departures, and account for them in a respecified model to be used as a basis for inferences pertaining to the primary hypotheses of interest.

Statistical adequacy refers to the validity–vis-à-vis data $\mathbf{z}_0$–of the probabilistic assumptions comprising the statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ in question, and provides *the* sole criterion for 'when $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ accounts for the (recurring) regularities in data $\mathbf{z}_0$.'

Error-reliability. Statistical adequacy renders the relevant error probabilities ascertainable by ensuring that the *nominal* error probabilities for assessing substantive claims are approximately equal to the *actual* ones. Despite its obvious importance, securing the reliability of inference has been largely neglected by the modern statistics literature because in addressing it one had to deal with two difficult hurdles. The *first* was to specify $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ explicitly using a complete set of testable – vis-a-vis data $\mathbf{z}_0$ – assumptions. The *second*, and more difficult, had to do with delineating the role of *substantive* subject matter information in specifying $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$; see Lehmann (1990). The crucial difficulty, which has devastated the trustworthiness of empirical modeling in the social sciences, arises when one imposes the substantive information (theory) on the data at the outset. The end result is often a statistically and substantively misspecified model, but one has no way to delineate between the two sources of error:

$$\boxed{\text{is the substantive information false? or are the inductive premises mispecified?}} \quad (3)$$

This is an example of a classic problem in philosophy of science known as Duhem's problem; see Mayo (1997). To place it in proper context let us briefly discuss how it has stumped econometric modeling since the early 20th century.

3.3 Statistical misspecification vs. the realisticness issue

The huge gap between economic theory concepts and the corresponding available data (problem [v]) has frustrated early attempts to give data a more substantive role by focusing on data-driven models. Such attempts are problematic mainly because they implicitly assume that the theory-concepts largely coincide with the selected data, and rely on goodness-of-fit measures to assess when a model 'accounts for the regularities' in the data ([iii]). These attempts had disastrous consequences for empirical modeling in economics because they inadvertently contributed to the entrenchment of the PET perspective in a chain of misleading arguments.

Statistical unreliability. Data-driven models in the social sciences, like correlation analysis, linear regression, factor analysis and principal component analysis, have been disastrously unreliable when applied to observational data, primarily because they relied on goodness-of-fit measures to assess model adequacy, instead of validating their underlying inductive premises.

Statistical spuriousness. In turn, the arbitrariness of goodness-of-fit measures created a strong impression that one can 'forge' significant correlations (or regression coefficients) at will, if one was prepared to persevere long enough 'mining' the data. This mistaken impression is almost universal among social scientists as well as philosophers. What they ignore is that when the underlying inductive premises are misspecified the apparent 'significance' is usually an artifact.

Misplaced role for substantive information. In turn, this erroneous impression has led to widely held (but fallacious) belief that substantive subject matter (theory) information provides the only safeguard against statistical spuriousness. In reality, statistical spuriousness can be easily detected using statistical M-S tests.

Exploiting these confusions the PET perspective consolidated its dominance on economic modeling and persistently charged any alternative perspective that took the data seriously as yet another form of 'measurement without theory', 'data-mining'and 'hunting' for statistical significance and the like.

Admonitions and berates concerning the devastating effects of using statistically misspecified models to draw inferences fall in deaf ears with the advocates of the PET perspective because to them such admonitions sound like a well-rehearsed complaint concerning the *unrealisticness* of their structural models, going back to Malthus (1836) who criticized the Ricardian method as based on 'premature generalization' which occasions "an unwillingness to bring their theories to the test of experience."(ibid, p. 8). Modern advocates of the PET perspective, respond by invoking the authority of Friedman (1953) to counter that such unrealisticness is inevitable, since all models are idealizations and not exact descriptions of reality; see Uskali (2009). This, however, exemplifies a major confusion between statistical and substantive inadequacy. There is a crucial difference between:

(a) the *unrealisticness* of the substantive assumptions comprising the theory-model (substantive premises), vis-à-vis the phenomenon of interest, and

(b) the *invalidity* of the probabilistic assumptions comprising the statistical model (inductive premises), vis-à-vis the data in question.

This is primarily because the types of errors one should probe for as well as guard against are very different in the two cases. Unfortunately, the PET perspective ignores this distinction and often foists the theory-model on the data at the outset giving rise to both *statistically* and *substantively misspecified* models. The only way to address the Duhemian ambiguity in (3) is to secure the statistical adequacy first in order to render reliable the statistical tools for assessing the substantive adequacy.

The realisticness of the theory is an issue that pertains to the substantive adequacy of the estimated model vis-à-vis the phenomenon of interest, i.e. whether the model in question provides a veritable explanation for that phenomenon. Substantive adequacy concerns the extent to which the estimated model accounts for all systematic aspects of the reality it aims to explain in a statistically and substantively adequate way, shedding light on the phenomenon of interest, i.e. 'learning from data'. Such inadequacy can easily arise from impractical *ceteris paribus* clauses, external invalidity issues, missing confounding factors, false causal claims, etc.; see Guala (2005). Securing substantive adequacy calls for additional probing of (potential) errors in bridging the gap between theory and data. However, without securing statistical adequacy first, such probing is likely to be misleading because the statistical procedures employed cannot be trusted to yield reliable inferences.

The PET perspective relies on certain statistical indicators, such as goodness-of-fit and goodness-of-prediction statistics, as well as several subjective judgements pertaining to the model's capacity to 'shed light' and/or confirm preconceived beliefs by the modeler. What is often ignored is that the invoked statistical "metrics of goodness" are, at best, questionable, without statistical adequacy. For example Friedman (1953), p. 8, argues:

> "Viewed as a body of substantive hypotheses, theory is to be judged by its predictive power for the class of phenomena which is intended to 'explain'."

This, however, begs the question:

▶ How can one reliably appraise the predictive power of a theory when the reliability of the very tools (statistical inference) used in that assessment is at best unknown, but potential highly misleading?

Friedman (p. 8) goes on to make a case for confronting theory with data:

> "Only factual evidence can show whether it is 'right' or 'wrong',

a claim which also begs a related question:

▶ How does one establish 'factual evidence' without statistical adequacy?

It is well-known that raw data do not constitute readily available (reliable) evidence one can use to confront theories with; see Spanos (2010a).

3.4 Addressing Duhemian ambiguities

The key to dealing with the crucial Duhemian ambiguity in (3) is to distinguish, *ab initio*, between *statistical* and *substantive* information and then bridge the gap between them by a sequence of interconnecting models which enable one to delineate and probe for the potential errors at different stages of modeling; Spanos (1986). From the theory side, the substantive information is initially encapsulated by a theory model and then modified into a structural one $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ to render it estimable with data $\mathbf{z}_0$. From the data side, the statistical information is distilled by a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ whose parameterization is chosen with a view to render $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$ a reparametrization/restriction thereof, say $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}$. The statistical adequacy of $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is secured first in order to ensure the reliability of the procedures for appraising *substantive* claims including the overidentifying restrictions arising from $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}$; formal assessment of the latter provides a way to reconcile the two sources of information; Spanos (1990, 2007a). The underlying rationale is that statistical adequacy is a precondition for securing the reliability of the inference procedures used in appraising substantive adequacy because error-reliability [the *actual* error probabilities approximate closely the *nominal* ones]. A statistically misspecified model will lead inductive inferences astray.

The big hurdle in getting a handle on the reliability of inference has been to establish a notion of 'statistical information' that can be untangled, at least ab initio, from substantive information.

A. A purely probabilistic construal of a statistical model

Spanos (1986) proposed a notion of *statistical information* that relates directly to the chance regularity patterns (distribution, dependence and heterogeneity) exhibited by data $\mathbf{z}_0$ when the latter is viewed as a realization of a *generic* – free from any substantive information – stochastic process $\{Z_k,\ k\in\mathbb{N}:=(1,2,...)\}$. This notion enables one to put forward a purely probabilistic construal of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ by viewing it as a particular *parameterization* of the probabilistic structure of a process $\{Z_k,\ k\in\mathbb{N}\}$. To notion of statistical information is illustrated in figures 1-4 where no substantive information pertaining to what these data measure is available, but one can observe certain chance regularities exhibited by these data series. One can easily distinguish between a typical realization of a NIID (fig. 1) vs. an Exponential

IID (fig. 2). Similarly, one can easily detect the difference between a typical realization of a Normal, Markov dependent and stationary process (fig. 3) from a NIID (fig. 1). Similarly, fig. 4 exhibits a typical realization of a Normal, Independent but mean-heterogeneous process; see Spanos (1999) ch. 5 for further details. The chance regularities exhibited by the data in figures 1-4 can be detected independently of any substantive information pertaining to the variables underlying these data.

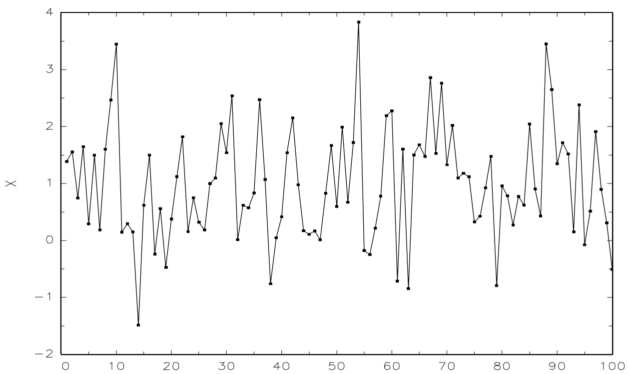

Fig. 1 - a typical realization of a NIID process

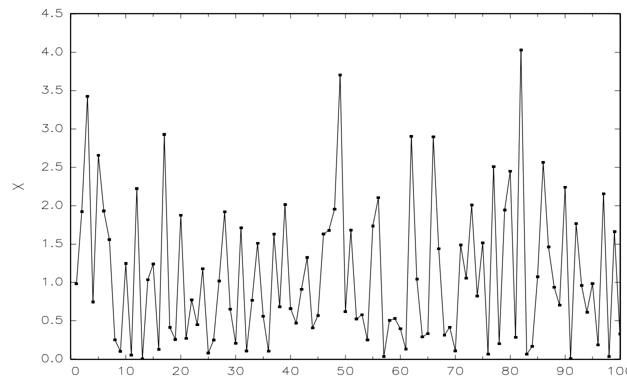

Fig. 2 - a typical realization of an Exponential IID process

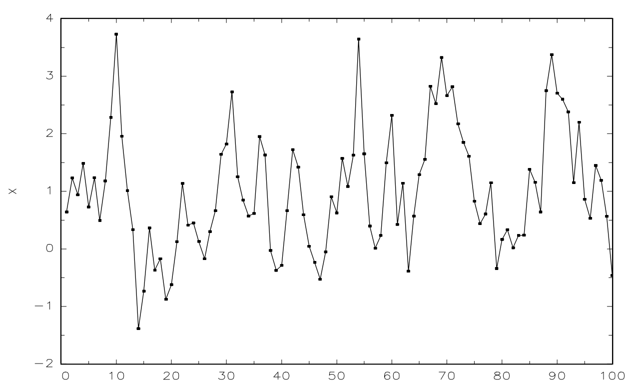

Fig. 3 - a typical realization of a Normal, Markov, Stationary process

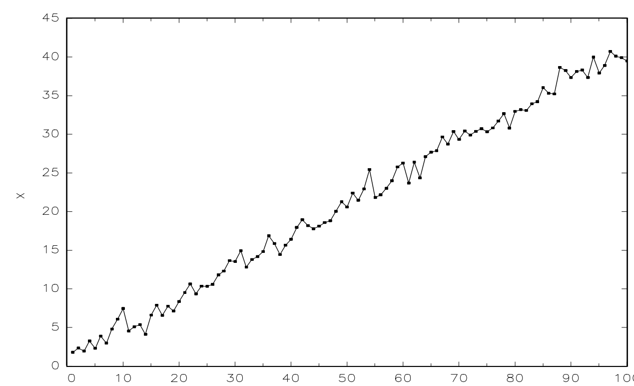

Fig. 4 - a typical realization of a NI, but mean-treding process

Substantive subject matter information usually enters empirical modeling in the form of a *structural model,* say $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{z})$, which constitutes an *estimable form* of a theory, in view of the specific data $\mathbf{z}_0$.

B. Reconciling substantive and statistical information

The statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ is built exclusively on *statistical systematic information* in data $\mathbf{z}_0$, and is selected so as to meet two interrelated aims:

(I) to account for the chance regularities in data $\mathbf{z}_0$ by choosing a probabilistic structure for the stochastic process $\{Z_k,\ k\in\mathbb{N}\}$ underlying $\mathbf{z}_0$ so as to render it a 'typical realization' thereof, and

(II) to parameterize the probabilistic structure of $\{Z_k,\ k\in\mathbb{N}\}$ in the form of an adequate statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ that would *embed* $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$ in its context, via *reparametrization/restriction* $\mathbf{G}(\boldsymbol{\theta}, \boldsymbol{\varphi})=\mathbf{0}$; formal assessment of the latter provides a way to reconcile the two sources of information; Spanos (1990, 2007a).

The Probabilistic Reduction (PR) approach (Spanos, 1989) provides the framework for addressing [iii]-[v] by:

(i) specifying $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ in terms of a *complete* list of (internally consistent) probabilistic assumptions, in a form that is testable vis-à-vis data $\mathbf{z}_0$, and

(ii) supplementing that with a *statistical generating mechanism* (GM) to provide a bridge between the statistical and substantive information.

Having a complete list of *testable* [vis-à-vis data $\mathbf{z}_0$] probabilistic assumptions, specifying a statistical model, often demands:

(a) unveiling *hidden assumptions,* as well as

(b) recasting assumptions about *unobservable errors* $\{\varepsilon_k,\ k\in\mathbb{N}\}$ as in the case of linear regression and related models, into equivalent assumptions concerning the *observable* stochastic processes $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$ involved; see Spanos (1986).

It is important to emphasize that Bayesian inference is equally, if not more, vulnerable to the statistical misspecification problem and implications thereof. This is because the likelihood function of a statistically misspecified model will undermine the reliability of the posterior distribution, irrespective of the appropriateness of the prior distribution.

4 Statistical vs. substantive premises of inference

In this section we discuss the problem of distinguishing between the statistical and substantive premises of inference, by bringing out the confusions between the two in traditional textbook treatments. To make the discussion less abstract we focus on the Linear Regression (LR), which is arguably the quintessential statistical model of interest in econometrics and other social sciences.

4.1 The traditional textbook Linear Regression model

The central role of the Linear Regression (LR) model can be surmised from the fact that econometrics textbooks treat most statistical models of interest in econometrics as variations/extensions of the LR model; see Greene (2008). Despite its pivotal role in econometric modeling, one will be hard pressed to find any two econometric textbooks, from the pioneering books of Johnston (1963) and Goldberger (1964), to the most current, Greene (2008), Wooldridge (2009) and Stock and Watson (2008) inter alia, which give the same specification (statistical assumptions) for the LR model. Moreover, a closer look at such specifications indicate several redundancies, ambiguities and confusions between statistical and substantive information.

In an effort to delineate some of these issues consider a typical specification of the LR model as given in Greene (2008) and Kennedy (2008), together with certain envisioned departures that elucidate the intended scope of the assumptions.

A1. Linearity: The model specifies a linear relationship between the variables y_k and $\mathbf{x}_k:=(x_{1k}, x_{2k}, ..., x_{mk})^\top$ of the form:

$$y_k = \beta_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_k + \varepsilon_k,\ k{=}1,2,...,n,$$

where $\boldsymbol{\beta}_1^\top{:=}(\beta_1, \beta_2, \cdots, \beta)$, with constant coefficients β_0 and $\boldsymbol{\beta}_1$.

Examples of departures:

(a) Wrong regressors (inclusion of irrelevant and omission of relevant variables).

(b) Nonlinearity ($y_k = h(x_{1k}, x_{2k}, ..., x_{mk}) + \varepsilon_k$, where $h(.)$ is a nonlinear function).

(c) Nonconstant coefficients (i.e. $\beta_0(k), \beta_1(k), \beta_2(k), ..., \beta_m(k)$ for $k{=}1,2,...,n$).

A2. Exogeneity: $E(\varepsilon_k \mid \mathbf{X}_k{=}\ \mathbf{x}_k) = 0$, for $k{=}1,2,...,n$.

Departure: $E(\mathbf{x}_k\varepsilon_k) \neq \mathbf{0}$, for $k{=}1,2,...,n$.

A3. Homoskedasticity: $E(\varepsilon_k^2 \mid \mathbf{X}_k{=}\ \mathbf{x}_k) = \sigma^2 < \infty$, for all $k{=}1,2,...,n$.

Departure: $E(\varepsilon_k^2 \mid \mathbf{X}_k{=}\mathbf{x}_k) = \sigma_k^2 < \infty$, for $k{=}1,2,...,n$.

A4. Nonautocorrelation: $E(\varepsilon_k\varepsilon_j \mid \mathbf{x}_k) = 0$, for $k \neq j,\ k,j{=}1,2,...,n$.

Departure: $E(\varepsilon_k\varepsilon_j \mid \mathbf{X}_k{=}\ \mathbf{x}_k) = \omega_{kj} \neq 0$, for $k \neq j,\ k,j{=}1,2,...,n$.

A5. Fixed in repeated samples: the observations on $\mathbf{x}_k$, $k{=}1,2,...,n$, are viewed as fixed in repeated samples.

Examples of departures:

(a) Errors in variables ($x_{ik}^* = x_{ik} + \epsilon_{ik},\ i{=}1,2,...,m$)

(b) Lagged y_k variables, say y_{k-1}, y_{k-2}, *among the regressors* $\mathbf{x}_k$.

(c) Simultaneity of certain regressors.

A6. Full rank: For $\mathbf{X}{:=}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)$, where $\mathbf{x}_i{:=}(x_{i1}, x_{i2}, ..., x_{in})^\top$, $i{=}1,2,...,m$:

$$\text{rank}(\mathbf{X}) = m,\ \text{for } n > m.$$

Departures: (a) rank$(\mathbf{X}) < m$, (b) $(\mathbf{X}^\top\mathbf{X})$ is ill-conditioned.

A7. Normality: $(\varepsilon_k \mid \mathbf{X}_k{=}\ \mathbf{x}_k) \backsim \mathsf{N}(.,.)$, for $k = 1,2,...,n$.

The most direct way to bring out the various redundancies, ambiguities and confusions in the above list of assumptions is to contrasting it to the list of the probabilistic assumptions specifying the LR model when viewed as a statistical model in purely probabilistic terms.

4.2 Statistical model: purely probabilistic construal

This probabilistic construal of a statistical model takes the choice of the (potentially) relevant variables $\mathbf{Z}_t{:=}(y_t, \mathbf{X}_t)$ and the corresponding data $\mathbf{Z}_0{:=}(\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_n)$ as given (selected on the basis of certain substantive information), and proceeds to give data $\mathbf{Z}_0$ 'a life of its own' by constructing the statistical model exclusively on the probabilistic information reflected in the chance regularity patterns exhibited by data $\mathbf{Z}_0$; see Spanos (1986). The specification of the statistical model is viewed as arising from choosing the probabilistic structure of the stochastic process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$, underlying data $\mathbf{Z}_0$, with the dual objective in mind: (a) to render $\mathbf{Z}_0$ a 'typical realization' thereof [a reformulation/extension of Fisher's "Of what population is this a random sample?"], and (b) to parametrize this structure in a way that can (potentially) encompass parametrically the substantive (structural) model of interest.

The specification of the Normal/Linear Regression (N/LR) model given in table 1 in terms of a statistical Generating Mechanism (GM) and the probabilistic assumptions [1]-[5], pertaining to the conditional distribution $D(y_t \mid \mathbf{X}_t;\boldsymbol{\theta})$, constitute a purely probabilistic rendering of this particular statistical model. The question that naturally arises is 'how does not come up with such a purely probabilistic specification?' Although the details will be discussed in the next subsection, it is important to bring out right away the fact that the assumptions [1]-[5] rely sole on $D(y_t \mid \mathbf{X}_t;\boldsymbol{\theta})$, and the specification of the statistical GM involves no substantive information, but provides the bridge for reconciling the two sources of information; see Spanos (2006c).

Table 1 - Normal/Linear Regression model

Statistical GM:	$y_t = \beta_0 + \boldsymbol{\beta}_1^{\top}\mathbf{x}_t + u_t, \quad t{\in}\mathbb{N}{:=}(1,2,..,n,...)$	
[1] Normality:	$(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t) \backsim \mathsf{N}(.,.),$	$t{\in}\mathbb{N}.$
[2] Linearity:	$E\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = \beta_0 + \boldsymbol{\beta}_1^{\top}\mathbf{x}_t,$	
[3] Homoskedasticity:	$Var\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = \sigma^2,$	
[4] Independence:	$\{(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$ indep. process,	
[5] t-invariance:	$\boldsymbol{\theta}{:=}(\beta_0,\boldsymbol{\beta}_1,\sigma^2)$ are *not* varying with t,	

$\beta_0{=}\left[\mu_1 - \boldsymbol{\beta}_1^{\top}\boldsymbol{\mu}_2\right]\in\mathbb{R},\ \boldsymbol{\beta}_1{=}\left[\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}\right]\in\mathbb{R},\ \sigma^2 = \left[\sigma_{11} - \boldsymbol{\sigma}_{21}^{\top}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}\right]\in\mathbb{R}_+,$
$\mu_1{=}E(y_t),\ \boldsymbol{\mu}_2{=}E(\mathbf{X}_t),\ \sigma_{11}{=}Var(y_t),\ \boldsymbol{\sigma}_{21}{=}Cov(\mathbf{X}_t,y_t),\ \boldsymbol{\Sigma}_{22}{=}Cov(\mathbf{X}_t).$

Indeed, the statistical GM is viewed as a purely probabilistic notion based on the orthogonal decomposition:

$$y_t = E\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) + u_t, \tag{4}$$

where by definition, $\mu_t = E\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right)$ is the *systematic* component, and $u_t = y_t - E\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right)$ the *non-systematic* component (error term term), and they satisfy the following properties:

$$\begin{array}{ll}\text{[i]} & E\left(u_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = E\left[y_t - E\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) \mid \mathbf{X}_t{=}\mathbf{x}_t\right] = 0,\\ \text{[ii]} & E\left(\mu_t u_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = \mu_t E\left(u_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = 0,\\ \text{[iii]} & E\left(u_t^2 \mid \mathbf{X}_t{=}\mathbf{x}_t\right) = Var\left(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t\right).\end{array} \tag{5}$$

The decomposition in (4) specifies a statistical GM describing an idealized stochastic mechanism pertaining to how data $(y_1, y_2, ...y_n)$ can be generated given the parameters $(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$ and the observations $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$. A more operational form of this statistical GM is:

$$y_t = \beta_0 + \boldsymbol{\beta}_1^{\top}\mathbf{x}_t + \sigma\varepsilon_t,\ \varepsilon_t \backsim \mathsf{N}(0,1), \quad t{=}1,2,..,n, \tag{6}$$

showing how one can generate artificial data for y_t using pseudo-random numbers ε_t from $\mathsf{N}(0,1)$.

To avoid a long digression on how the PR specification of the N/LR model gives rise to the model as given in table 1, we postpone that discussion after the comparison of assumptions [1]-[5] with the textbook specification A1-A7.

4.3 Redundancies, ambiguities and confusions

4.3.1 The nature of the error term

The first thing to note before comparing the two lists of assumptions A1-A7 and [1]-[5] is the nature of the two error terms. In the textbook specification, the error $\varepsilon_k = (y_k - \beta_0 - \boldsymbol{\beta}_1^\top \mathbf{x}_k)$, is viewed as an *autonomous* term that represents all unmodeled influences on y_k. In this sense, the assumptions pertaining to the error process $\{(\varepsilon_k \mid \mathbf{X}_k{=}\,\mathbf{x}_k),\ k{\in}\mathbb{N}\}$ can be negated by both statistical and substantive misspecifications.

In contrast, as shown in (5), the statistical error $u_t = y_t - \beta_0 - \boldsymbol{\beta}_1^\top \mathbf{x}_t$ (table 1) is a *derived* term whose probabilistic structure is determined by that of the observable process $\{\mathbf{Z}_t,\ t{\in}\mathbb{N}\}$. Hence, the prespecified behavior of the error process $\{(u_t \mid \mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$ is belied only by statistical misspecifications.

A comparison between assumptions A1-A7 and [1]-[5] in table 1 reveals that the only pair of directly equivalent assumptions is [1] and A7. Assumption [2] is mathematically equivalent to the traditional exogeneity assumption A2 since:

$$E(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t) = \beta_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t \ \Leftrightarrow E(u_t \mid \mathbf{X}_t{=}\mathbf{x}_t) = 0,\ \text{for all } t{\in}\mathbb{N}.$$

In light of this, assumption [2] renders the linearity part of assumption A1 redundant. A closer look at assumptions A1-A2 reveals that they suffer from serious ambiguities because the linearity in A1 and some of the envisaged departures from A2, like omitted variables and simultaneity ($E(\mathbf{x}_k \varepsilon_k) \neq \mathbf{0}$), are clearly motivated by substantive information because it pertains to the adequacy of the theoretical explanation. Moreover, it is important to stress that in a number of textbooks the assumption $E(\mathbf{x}_k \varepsilon_k) = \mathbf{0}$ is treated as distinct from A2 (see Berry, 1993, Gujarati, 2009 inter alia), not realizing that mathematically $E(\varepsilon_k \mid \mathbf{X}_k{=}\,\mathbf{x}_k) = 0 \Rightarrow E(\mathbf{x}_k \varepsilon_k) = \mathbf{0}$. The remaining part of A1, relating to the constancy of the coefficients, in conjunction with assumption A3, are equivalent to assumption [5].

Assumption A3 suffers from a serious ambiguity in the sense that the term homoskedasticity, coined by Pearson 1905, constitutes a misnomer in this case because the perceived departure conflates two different problems:

(a) heterogeneity: $E(\varepsilon_k^2 \mid \mathbf{X}_k{=}\mathbf{x}_k){=}\sigma_k^2,\ k{\in}\mathbb{N}$,

(b) heteroskedasticity: $E(\varepsilon_k^2 \mid \mathbf{X}_k{=}\mathbf{x}_k){=}h(\mathbf{x}_k),\ k{\in}\mathbb{N}$,

the key difference being that (b) denotes the conditional variance being a function of $\mathbf{x}_k$, not k. The distinction is crucially important, because, as one be seen from table 2, the former is due to departures from the ID, but the latter from the Normality assumption. Conflating (a) and (b) will misdirect any attempts to respecify the statistical model with a view to secure statistical adequacy; see Spanos (1986).

There is a direct relationship between A4 and [4], in so far as the latter is generally a stronger assumption, even though the two assumptions are equivalent in the context of a Normal process. In a welcome recent trend in textbook econometrics, A4 has been replaced by the independence assumption concerning the observable processes $\{(y_t, \mathbf{X}_t),\ t{\in}\mathbb{N}\}$; see Spanos (1986) for an early call in that direction. Indeed, Stock and Watson (2007) and Wooldridge (2009) replace A4 with a more restrictive *random sample* assumption:

A4*. **Random sampling**: $(\mathbf{X}_t, y_t),\ t=1,2,...,n$ are independently and identically distributed (IID) draws from their joint distribution.

Although sufficient for [4], as indicated by the reduction (table 2), assumption A* is clearly not necessary.

Assumption A5 is obviously related to all the other assumptions A1-A4 and A6-A7 because it concerns the conditioning variables $\mathbf{X}_t$, stating that the only information pertaining to $\mathbf{X}_t$ considered relevant in the context of the LR model are the particular observed values $\{(\mathbf{X}_t{=}\mathbf{x}_t),\ t=1,2,...,n\}$, and not the sigma-field generated by $\mathbf{X}_t$ $(\sigma(\mathbf{X}_t))$. The conditions under which the only relevant information for any inference pertaining to the parametrization of interest $\boldsymbol{\theta}{:=}(\beta_0, \boldsymbol{\beta}_1, \sigma^2)$, have been formalized in the form of weak exogeneity proposed by Engle et al (1983), which will be discussed in the next subsection. It turns out that in the context of the specification [1]-[5], the explicit parametrization for $\boldsymbol{\theta}$ given in table 1 renders the weak exogeneity assumption redundant. The key restriction pertaining to A5 is the use a *single* realization of the process $\{\mathbf{X}_t,\ t{\in}\mathbb{N}\}$ to generate artificial realizations of the $\{(y_t|\ \mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$ process using the statistical GM (6) as follows.

Step 1: Specify the 'true' values of (or estimate) the unknown parameters, say $\boldsymbol{\theta}^*{:=}(\beta_0^*, \boldsymbol{\beta}_1^*, \sigma_*^2)$.

Step 2: Generate N realizations of size n of $\varepsilon_t \backsim \mathsf{N}(0,1)$: $\left(\boldsymbol{\varepsilon}^{(1)}, \boldsymbol{\varepsilon}^{(2)}, \cdots, \boldsymbol{\varepsilon}^{(N)}\right)$, where $\boldsymbol{\varepsilon}^{(k)} := (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^\top$.

Step 3: Use the observed values of $\{(\mathbf{X}_t{=}\mathbf{x}_t),\ t{=}1,2,...,n\}$, $\mathbf{X}_0{:=}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m)$, where each $\mathbf{x}_k$ represents n observations on the k-th variable.

Step 4: Substitute sequentially each $\boldsymbol{\varepsilon}^{(k)}$ and $\mathbf{X}_0$ into the statistical GM:

$$\mathbf{y}^{(k)} = \mathbf{1}\beta_0^* + \mathbf{X}_0\boldsymbol{\beta}_1^* + \sigma_*\boldsymbol{\varepsilon}^{(k)}, \quad k{\in}\mathbb{N}, \tag{7}$$

to generate the artificial realizations: $\left(\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(N)}\right)$, where $\mathbf{y}^{(k)}{:=}(y_1, ..., y_n)^\top$.

These 'artificial' data realizations can then be used to operationalize the sampling distributions of any estimators, test statistics and predictors in the context of the N/LR model one might be interested in. To bring out the essence of A5 one needs to contast the above simulation with an alternative one where one uses a different realization drawn from the process $\{\mathbf{X}_t,\ t{\in}\mathbb{N}\}$ at each iteration:

Step 3*: Generate N *realizations* of the IID process $\{\mathbf{X}_t,\ t{\in}\mathbb{N}\}$ from $\mathbf{X}_t \backsim \mathsf{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$, say $\left(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots \mathbf{X}^{(N)}\right)$.

Step 4*: Substitute sequentially each $\boldsymbol{\varepsilon}^{(k)}$ and $\mathbf{X}^{(k)}$ into the statistical GM:

$$\mathbf{y}^{(k)} = \mathbf{1}\beta_0^* + \mathbf{X}^{(k)}\boldsymbol{\beta}_1^* + \sigma_*\boldsymbol{\varepsilon}^{(k)}, \quad k{\in}\mathbb{N}, \tag{8}$$

to generate the 'artificial' realizations: $\left(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(N)}\right)$. Steps 3*-4* will generate data associated with the process $\{(y_t|\sigma(\mathbf{X}_t)),\ t{\in}\mathbb{N}\}$ as opposed to $\{(y_t|\ \mathbf{X}_t{=}\mathbf{x}_t),\ t{\in}\mathbb{N}\}$.

Looking at the perceived departures (a)-(c) from A5, it becomes clear that they constitute a mixture of both statistical and substantive misspecifications since errors-of-measurement and simultaneity issues are raised when the above model is supplemented with *additional* (substantive) information pertaining to how the variables in $\mathbf{X}_t$ (or a subset) were generated. The weak exogeneity conditions (i)-(ii) can be repealed in two different ways (Spanos, 2007a). The first is by way of statistical information such as the distribution of $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is Student's t instead of Normal. The second is via substantive information which imposes over-identifying restrictions among the statistical parameters $\boldsymbol{\theta}$. Moreover, it is important to emphasize that that when $\mathbf{X}_t$ is treated as a random vector, i.e. the conditioning is with respect to $\sigma(\mathbf{X}_t)$ (not $\mathbf{X}_t$=$\mathbf{x}_t$), the statistical GM changes to:

$$y_{\ t} = E(y_t \mid \sigma(\mathbf{X}_t)) + u_t = \beta_0 + \boldsymbol{\beta}_1^\top \mathbf{X}_t + \sigma\varepsilon_t,\ \ \varepsilon_t \backsim \mathsf{N}(0,1),\quad t{=}1,2,..,n, \tag{9}$$

but no violations of assumptions [1]-[5] are implied; see Spanos (1986) for the details. Indeed, all three conditions, [i]-[iii] are still valid for a stochastic $\mathbf{X}_t$, including the so-called exogeneity assumption A2, in the sense that for $u_t = y_t - E(y_t \mid \sigma(\mathbf{X}_t))$:

$$\begin{array}{rl}
\text{[i]} & E\left(u_t \mid \sigma(\mathbf{X}_t)\right) = E\left[y_t - E\left(y_t \mid \sigma(\mathbf{X}_t)\right)|\sigma(\mathbf{X}_t)\right] = 0, \\
\text{[ii]} & E\left(\mu_t u_t \mid \sigma(\mathbf{X}_t)\right) = \mu_t E\left(u_t \mid \sigma(\mathbf{X}_t)\right) = 0, \\
\text{[iii]} & E\left(u_t^2 \mid \sigma(\mathbf{X}_t)\right) = Var\left(y_t \mid \sigma(\mathbf{X}_t)\right).
\end{array}$$

The full rank assumption A6 is not a probabilistic assumption as such, but it concerns the minimal adequateness of the observed information in data $\mathbf{X}$. Hence, unless there are accounting identities among the columns of $\mathbf{X}$, it must be of full rank. If there is not enough variability in the observed data $\mathbf{X}$ the real problem is likely to be the ill-conditioning of the $(\mathbf{X}^\top\mathbf{X})$ matrix. In the econometric literature the two problems (a) and (b) are often confused; see Spanos and McGuirk (2002).

The most problematic among the assumptions A1-A7 is A1 because, in addition to the overlap with other assumptions, the perceived departures conflate substantive with statistical misspecifications. The non-linearity of the regression function and the non-constancy of the coefficients $(\beta_0, \boldsymbol{\beta}_1)$ are clearly statistical misspecifications associated with [2] and [5], respectively, but the wrong regressors problem is obviously a substantive misspecification issue. The real conundrum with mixing up statistical and substantive assumptions is that statistical adequacy is a *sine qua non* for assessing substantive adequacy because without the former one cannot appraise reliably substantive questions of interest. This stems from the fact that without securing statistical adequacy first, the reliability of the inference procedures pertaining to the substantive questions of interest is totally unknown; rendering such inferences at best questionable and at worst out-and-out erroneous. Statistical adequacy secures the reliability of such inferences, rendering them ascertainable by ensuring that the actual closely approximate the nominal error probabilities. The traditional textbook discussion of the omitted variables problem is conflating the statistical and substantive issues giving rise to an account revolving around estimation bias that makes no statistical or substantive sense; see Spanos (2006b).

Of particular interest in this paper is assumption [5] which constitutes one of the most crucial assumptions underlying empirical modeling in econometrics. This assumption has a long history in econometrics because Keynes (1939) argued with conviction that such an assumption is unlikely to be valid when modeling with economic data. His primary argument was that the real economy is inherently heterogenous over time. What his argument ignored was the fact that this probabilistic assumption is testable vis-à-vis data $\mathbf{Z}_0$! It is clear from the probabilistic reduction in (11) than when the process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ is non-ID (heterogeneous) in the sense that:

$$\mathbf{Z}_t := \begin{pmatrix} y_k \\ \mathbf{X}_k \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1(t) \\ \boldsymbol{\mu}_2(t) \end{pmatrix}, \begin{pmatrix} \sigma_{11}(t) & \boldsymbol{\sigma}_{21}^\top(t) \\ \boldsymbol{\sigma}_{21}(t) & \boldsymbol{\Sigma}_{22}(t) \end{pmatrix} \right), \tag{10}$$

the resulting N/LR model has *t-varying* parameters:

$$\beta_0(t) = \left[\mu_1(t) \text{-} \boldsymbol{\beta}_1^\top(t)\boldsymbol{\mu}_2\right],\ \boldsymbol{\beta}_1(t) = \left[\boldsymbol{\Sigma}_{22}^{-1}(t)\boldsymbol{\sigma}_{21}(t)\right],\ \sigma^2(t) = \left[\sigma_{11}(t) \text{-} \boldsymbol{\sigma}_{21}^\top(t)\boldsymbol{\Sigma}_{22}^{-1}(t)\boldsymbol{\sigma}_{21}(t)\right],$$

which cannot be estimated without imposing restrictions concerning the nature of the heterogeneity. One of the first attempts to impose such restrictions on $\boldsymbol{\beta}_1(t)$ was Swamy (1970, 1971) proposing to view it as random with constant mean and variance:

$$\boldsymbol{\beta}_1(t) = \boldsymbol{\beta}_1 + \mathbf{v}(t),\ v(t) \backsim \mathsf{IID}(\mathbf{0}, \mathbf{V}),\ \mathbf{V} = \text{diag}(v_1, v_2, ..., v_m) > \mathbf{0}.$$

Testing assumption [5] using a *single realization* of a non-ID process is often more challenging than the other assumptions because the generic alternative hypothesis, $\boldsymbol{\theta}(t) := (\beta_0(t), \boldsymbol{\beta}_1(t), \sigma^2(t))$, $t=1, 2, ..., n$, raises the *incidental parameter* problem - the number of parameters increases with the sample size. This explains why most M-S tests for t-invariance based on recursive estimates have very low power; see Kramer and Sonnberger (1986). To overcome this difficulty Koutris et al (2008) proposed to enhance the data information by generating N 'faithful replicas' of the original data $\mathbf{Z}_0$ using a Maximum Entropy (ME) resampling scheme.

4.4 The Probabilistic Reduction N/LR model specification

This probabilistic construal aims to separate the statistical from the substantive information and specify the statistical model exclusively on probabilistic information pertaining to the stochastic process $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$ underlying data $\mathbf{Z}_0$. Such a construal enables one to narrow down the set of all possible models $\mathcal{P}(\mathbf{z})$ that could have given rise to data $\mathbf{Z}_0$ to a single (family of) model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ using a *three-way* partitioning based on probabilistic assumptions pertaining to the process $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$.

DISTRIBUTION	DEPENDENCE	HETEROGENEITY
Normal	Independence	Identically Distributed
Beta	Correlation	Strict Stationarity
Gamma	Markov depend.	Weak Stationarity
Bernoulli	Long-memory	Separable heterogeneity
⋮	⋮	⋮

Example. The Normal/Linear Regression model (table 1) arises from partitioning $\mathcal{P}(\mathbf{z})$ by assuming that $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$ is assumed to be Normal (N), Independent (I) and Identically Distributed (ID) (NIID); see figure 5. The N/LR model constitutes a parameterization of such a process relying exclusively the conditional distribution $D(y_t \mid \mathbf{X}_t; \boldsymbol{\theta})$. The formal probabilistic reduction:

$$D(\mathbf{Z}_1,\mathbf{Z}_2,...\mathbf{Z}_n;\boldsymbol{\phi}) \rightsquigarrow \prod_{t=1}^{n} D(y_t \mid \mathbf{X}_t;\boldsymbol{\varphi}_1)$$

takes the following explicit form:

$$\begin{aligned} D(\mathbf{Z}_1,\mathbf{Z}_2,...\mathbf{Z}_n;\boldsymbol{\phi}) & \overset{\text{I}}{=} \prod_{t=1}^{n} D_t(\mathbf{Z}_t;\boldsymbol{\varphi}_t) \overset{\text{IID}}{=} \prod_{t=1}^{n} D(\mathbf{Z}_t;\boldsymbol{\varphi}) \overset{\text{IID}}{=} \\ & = \prod_{t=1}^{n} D(y_t \mid \mathbf{X}_t;\boldsymbol{\varphi}_1)\cdot D(\mathbf{X}_t;\boldsymbol{\varphi}_2), \ \ \mathbf{z}\in\mathbb{R}^{n(m+1)}. \end{aligned} \tag{11}$$

In the case where $D(\mathbf{Z}_t;\boldsymbol{\varphi})$ is multivariate Normal, i.e.

$$\mathbf{Z}_t := \begin{pmatrix} y_k \\ \mathbf{X}_k \end{pmatrix} \backsim \mathsf{N}\left(\begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{21}^\top \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right), \tag{12}$$

the conditional $D(y_t \mid \mathbf{X}_t;\boldsymbol{\varphi}_1)$ and marginal $D(\mathbf{X}_t;\boldsymbol{\varphi}_2)$ distributions are:

$$(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t) \backsim \mathsf{N}\left(\beta_0 + \boldsymbol{\beta}_1^\top \mathbf{x}_t,\ \sigma^2\right), \quad \mathbf{X}_t \backsim \mathsf{N}\left(\boldsymbol{\mu}_2,\ \boldsymbol{\Sigma}_{22}\right).$$

What enables one to ignore the marginal distribution $D(\mathbf{X}_t;\boldsymbol{\varphi}_2)$ is that $\mathbf{X}_t$ is ***weakly exogenous*** (Engle et al, 1983) with respect to $\boldsymbol{\theta}{:=}(\beta_0,\boldsymbol{\beta}_1,\sigma^2)$, in the sense that:

(i) $\boldsymbol{\varphi}_1$ and $\boldsymbol{\varphi}_2$ are *variation free* (independent), and

(ii) the parameters of interest are $\boldsymbol{\theta} = \boldsymbol{\varphi}_1$.

Conditions (i)-(ii) render $D(\mathbf{X}_t;\boldsymbol{\varphi}_2)$ irrelevant for inferences relating to $\boldsymbol{\theta}$.

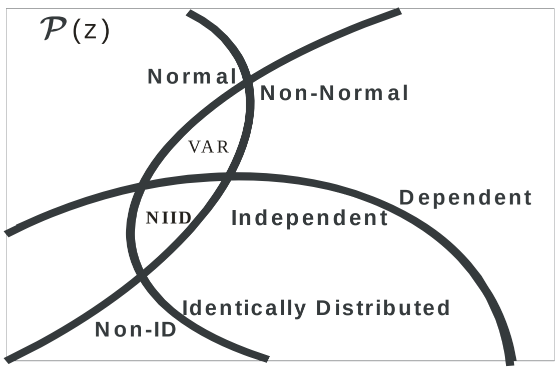

Fig. 5 - Statistical model specification by partitioning

Reduction vs. model assumptions. The above probabilistic reduction brings out the relationship between the reduction assumptions, NIID, pertaining to $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$, and the model assumptions [1]-[5]; see table 2. Broadly speaking, the former are sufficient for the latter, but often not necessary; see Spanos (1995) for necessary and sufficient conditions. This relationship is particularly useful for a number of different purposes, including specification, Mis-Specification (M-S) testing and respecification, because the NIID assumptions can be assessed using simple graphical techniques such as t-plots and scatter plots. Moreover, the implications of any ostensible departures from the NIID assumptions can often be traced to particular departures from model assumptions [1]-[5].

Table 2: Probabilistic assumptions

Reduction: $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$		Model: $\{(y_t \mid \mathbf{X}_t{=}\mathbf{x}_t),\ t\in\mathbb{N}\}$
N	$\longrightarrow$	[1]-[3]
I	$\longrightarrow$	[4]
ID	$\longrightarrow$	[5]

What is important for our purposes is that the model assumptions relate only to the probabilistic structure of $\{\mathbf{Z}_k,\ k\in\mathbb{N}\}$ as reflected in data $\mathbf{Z}_0$, without invoking any substantive information. For example, the t-plots in figures 2-3 indicate that the data underlying $\mathbf{Z}_t$ can be realistically viewed as typical realizations of a NIID process, irrespective of what these data series represent; see Spanos (1999), ch. 5.

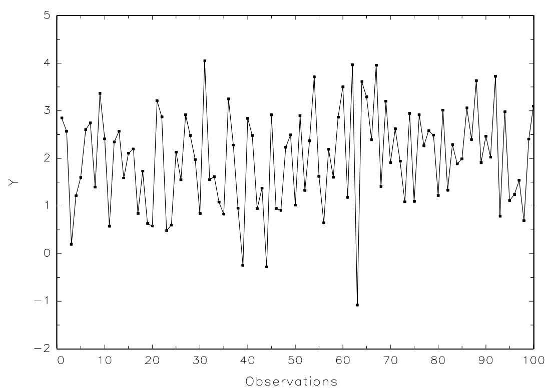

Fig. 6: t-plot of y_t

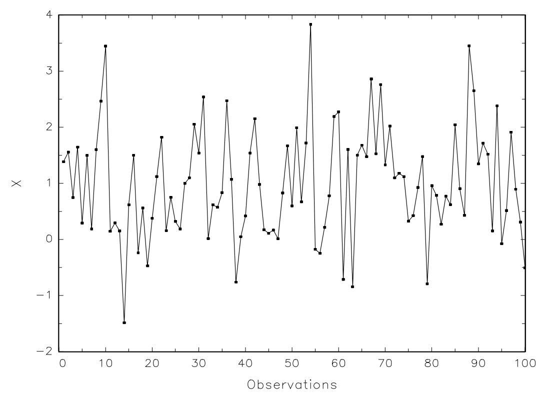

Fig. 7: t-plot of x_t

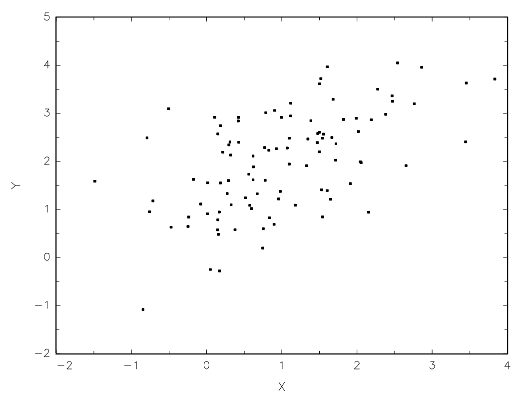

Fig. 8: Scatter-plot (x_t, y_t)

In addition, the scatterplot in figure 4 exhibits the elliptical shape associated with the bivariate Normal. Taken together the N/LR model would not be an unreasonable description of the statistical information reflected in figures 2-4. In this sense, the Normality, linearity and homoskedasticity assumptions [1]-[3] stems exclusively from the NIID assumption of $\mathbf{Z}_t{:=}(y_t, \mathbf{X}_t)$ in the sense that they hold irrespective of any substantive information pertaining to any linear structural relationship between the variables y_k and $\mathbf{X}_k$. In this sense, the model assumptions [1]-[5] are empirically based and thus directly testable vis-à-vis data $\mathbf{Z}_0$.

In contrast, assumptions [1]-[3] are unlikely to be satisfied by the data in figures 5-7 because, even though the exhibit IID features, all three figures strongly indicate that the underlying distributions are skewed, and thus, any theory claiming a linear structural relationship between y_k and x_k is likely to be falsified on empirical grounds. This is because the functional forms of the regression and skedastic functions for an asymmetric distribution like the Exponential are respectively:

$$E(y_t|X_t{=}x_t){=}\frac{1{+}\theta{+}\theta x_t}{(1{+}\theta x_t)^2},\quad Var(y_t|X_t{=}x_t){=}\frac{(1{+}\theta{+}\theta x_t)^2{-}2\theta^2}{(1{+}\theta x_t)^4},\ \text{for}\ x_t{\in}\mathbb{R}_+, \tag{13}$$

which differ drastically from those in assumptions [2]-[3]; Spanos (1999), ch. 7.

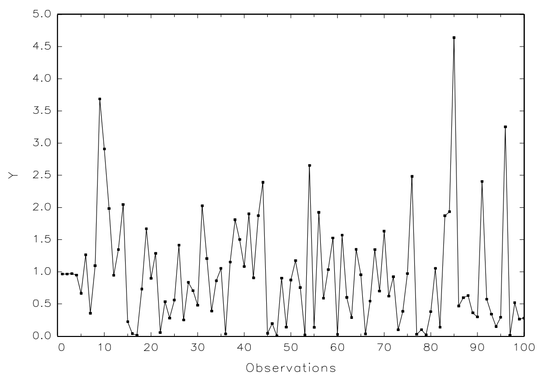

Fig. 9: t-plot of y_t

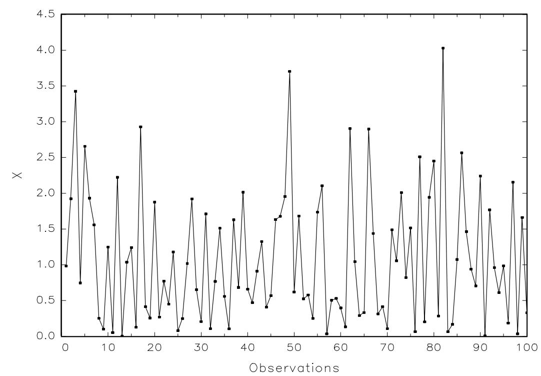

Fig. 10: t-plot of x_t

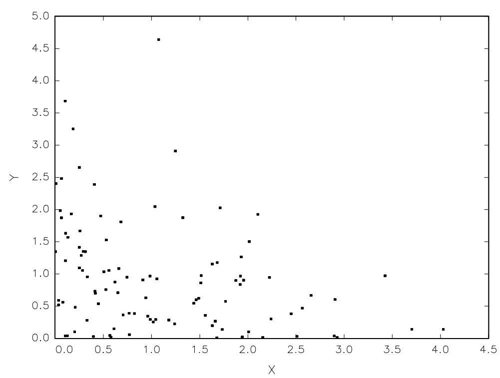

Fig. 11: Scatter-plot (x_t, y_t)

The use of data plots for detecting chance regularity patterns can render the various stages of statistical modeling, specification, M-S testing and respecification, considerably more effective. This is an important part of Exploratory Data Analysis (EDA), based on graphical techniques, in conjunction with analogical reasoning. To render the probing of M-S testing more effective, the statistician can use EDA to appraise the presence of potential violations before even selecting the relevant M-S tests. Far from being illegitimate 'data snooping' or 'data mining', graphical techniques provide powerful ways to select the type of M-S tests to apply to check assumptions most severely; see Spanos (2000), Mayo and Spanos (2004).

5 Revisiting the Keynesian Consumption function

In an attempt to shed light on several issues raised in the discussion of the previous three sections, let us consider a simple empirical example stemming from the substantive model based on Keynes' *Absolute Income Hypothesis* (AIH):

$$C = \alpha + \beta Y^D,\ \alpha > 0,\ 0 < \beta < 1. \tag{14}$$

The textbook approach suggests transforming this into a statistical model by attaching a Normal, white-noise error $\{u_t,\ t\in\mathbb{N}\}$:

$$C_t = \alpha + \beta Y_t^D + u_t,\ u_t \backsim \mathsf{NIID}(0, \sigma^2),\ t{=}1, 2, ..., n. \tag{15}$$

The statistical premises are defined in terms of the implicit *statistical model*, taking the form of the Normal/Linear Regression (table 1). The relevant data $\mathbf{z}_0{:=}\{(y_1, x_1), (y_2, x_2), ..., (y_n, x_n)\}$, represent annual USA time series data for the period 1947-1998, where $C_t{:=}\ y_t$ - real consumer's expenditure and $Y_t^D{:=}x_t$ - real personal disposable income. Estimating (15) by OLS yields:

$$y_t{=}\underset{(16.930)}{45.279} + \underset{(.007)}{.936}x_t + \widehat{u}_t,\ R^2{=}.997,\ s{=}49.422,\ n{=}52. \tag{16}$$

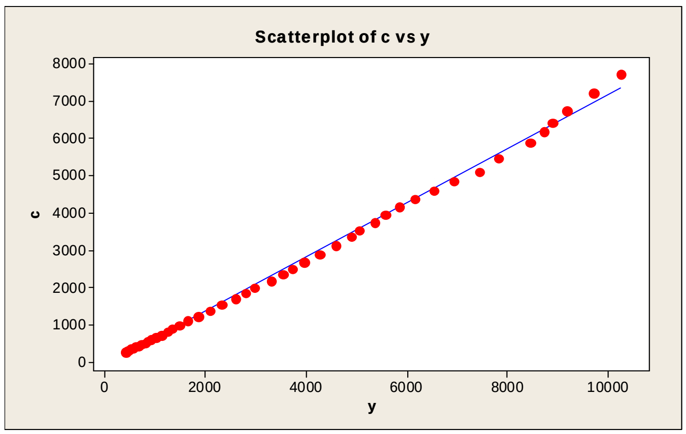

Fig. 12: Scatter-plot and fitted line

Taking these results at face value, it looks as though the goodness of fit ($R^2{=}.997$) is 'excellent' (see fig. 8) and the coefficients appear to be 'highly significant':

$$\tau_0(y){=}\tfrac{45.279}{16.930}{=}2.675[.004],\quad \tau_1(y){=}\tfrac{.936}{.007}{=}133.71[.000];$$

p-values in square brackets. Moreover, from the substantive perspective the estimated coefficients seem to have the correct signs and magnitudes ($\alpha{>}0,\ 0{<}\beta{<}1$).

Does the estimated model (16) provide *evidence for* the AIH? No! Most of the probabilistic assumptions of the N/Linear Regression model are invalid for this data. Using certain simple M-S tests (see Spanos and McGuirk, 2001), assumptions [2]-[5] are shown to be invalid as indicated by the small p-values in table 3.

Table 3 - Mis-Specification (M-S) tests for (16)	
Non-Normality:	$D'AP = 2.837[.242]$
Non-linearity:	$F(1, 49) = 154.822[.000]^*$
Heteroskedasticity:	$F(2, 47) = 7.344[.002]^*$
Autocorrelation:	$F(1, 48) = 159.274[.000]^*$
t-heterogeneity:	$F(1, 49) = 11.44[.000]^*$

The estimated model is *statistically misspecified*, indicating that the *actual error probabilities* are likely to be very different from their assumed *nominal* ones, voiding the above claims concerning the magnitude, sign and significance of the estimated coefficients! Worse, the estimated model in (16) is also likely to be substantively inadequate; confounding variables, etc. However, no question of substantive adequacy can be posed in the context of this statistically inadequate model as it stands. One needs to respecify and secure statistical adequacy before any substantive questions of interest can be appraised reliably.

Wild goose chases in research. What often happens in economics and the other social sciences when such an estimated model is published is something worth noting. Instead of other researchers questioning its statistical credentials and dismissing its findings as statistical artifacts, such a publication often creates a cottage industry of "smart" academics who provide "ingenious substantive explanations" why the coefficient of x_t, known as the marginal propensity to consume (MPC), is much higher than theory suggests it should be, in light of the fact that it gives rise to a large *multiplier effect* of $\frac{1}{1-.936}=15.625$. One paper after another are likely to make highly hyperbolized arguments on how to "explain away" the magnitude of the estimated coefficient β_0 using a variety of well-rehearsed sermons based on simultaneity, errors-in-variables, omitted variables, autocorrelation-corrected GLS estimates, etc., etc. When in fact, there is *nothing* to explain away because the statistical misspecification renders the above estimated coefficients, their standard errors, the t-ratios and the R^2 statistically meaningless numbers.

5.1 Ad hoc traditional textbook error-fixing

In order to illustrate how the traditional textbook approach is likely to make matters worse when a certain form os misspecification is detected, consider the most widely used Mis-Specification (M-S) test for *autocorrelation*, the Durbin-Watson (DW) test. This M-S test is based on comparing the original (M_0) to an alternative encompassing model (M_1):

$$\begin{array}{ll} M_0: & y_t=\beta_0+\beta_1 x_t+\varepsilon_t, \\ M_1: & y_t=\beta_0+\beta_1 x_t+u_t, \quad u_t=\rho u_{t-1}+\varepsilon_t, \end{array} \tag{17}$$

where ε_t is Normal, white noise, with the hypotheses of interest being:

$$H_0: \rho=0, \quad \text{vs.} \quad H_1: \rho \neq 0.$$

In the case of (16), the DW test rejects the null hypothesis since $DW=1.115[.021]$. What is a traditional modeler supposed to do next to address the detected misspecification? Adopt M_1 as the new respefied model and estimate it using Generalized Least Squares (GLS) to yield:

$$\begin{array}{ll} y_t=\underset{(.939)}{167.209}+\underset{(.037)}{1.898}x_t+\widehat{u}_t & u_t=\underset{(.152)}{.431}u_{t-1}+\widehat{\varepsilon}_t, \\ R^2=.996, \ DW=1.831, & s=1.742, \ n=35. \end{array} \tag{18}$$

The conventional wisdom (see Greene, 2008) considers the estimated model (18) as providing strong evidence for the encompassing model M_1.

This textbook strategy is a classic example of the *fallacy of rejection*, where evidence against H_0 is misinterpreted (the presence of temporal dependence) as evidence *for* the particular model underlying the alternative H_1. As argued in Spanos (2000), in no circumstances one adopts the alternative in a M-S test because the test in question has not probed all the different ways H_0 can be misspecified. H_1 assumes a particular violation of independence, and thus a rejection of the H_0 is evidence for the presence of autocorrelation, i.e. $\rho \neq 0$, but this by itself is not good evidence for M_1 – there was no chance to have uncovered the various other forms of dependence that would also give rise to $\rho \neq 0$. This is the kind of pejorative use of the data to construct (ad hoc) a model to account for an anomaly that critics often have in mind. The ways M_1 can be incorrect vis-a-vis data $\mathbf{z}_0$ have not been probed – M_1 has certainly *not passed a severe test*; see Mayo and Spanos (2004). This can be achieved by testing the assumptions of M_1. In the above case M_1 can be shown to be statistically *misspecified*.

A closer look at the traditional approach reveals that it does *not* provide systematic guidance for:

(a) an exhaustively complete probing strategy for *M-S testing*, or

(b) satisfactory answers to the *respecification* question.

As demonstrated in Spanos and McGuirk (2001), most of the textbook error-fixing strategies make matters worse, not better. As argued above, the Probabilistic Reduction approach was proposed as an alternative to the traditional textbook approach with a view to address the crucial methodological issues raised above.

5.2 The Probabilistic Reduction specification

In light of the annual USA time series data for the period 1947-1998, a Probabilistic Reduction (PR) econometrician will consider the appropriateness of the N/LR model (table 1) even before any inference is attempted. In particular, one can informally (using analogical reasoning) assess the appropriateness of the N/LR model assumptions [1]-[5], by assessing the reduction assumptions NIID in table 2, by posing the question: are the reduction assumptions:

(D) **Distribution**	(M) **Dependence**	(H) **Heterogeneity**
Normal?	Independent?	Identically Distributed?

appropriate for the consumption function data in figures 13-14? No!

These figures indicate most clearly that the IID assumptions are invalid; they exhibit both mean-trends and temporal dependence in the form of cycles; see Spanos (1999), ch. 5. Indeed, to bring out the latter more clearly, one can de-trend the data by regressing them on a trend polynomial (say order 2) and taking the residuals shown in figures 4-5. This corresponds to the philosopher's counterfactual reasoning!

The de-trended (D) data shown in figures 15-16 bring out the cycles more clearly and the educated guesses score card now reads:

(D) **Distribution**	(M) **Dependence**	(H) **Heterogeneity**
Normal?	non-Independent	mean-heterogeneous

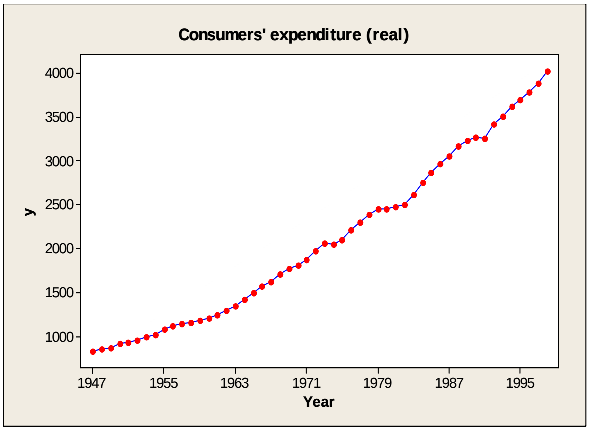

Fig. 13: t-plot of consumers' expenditure

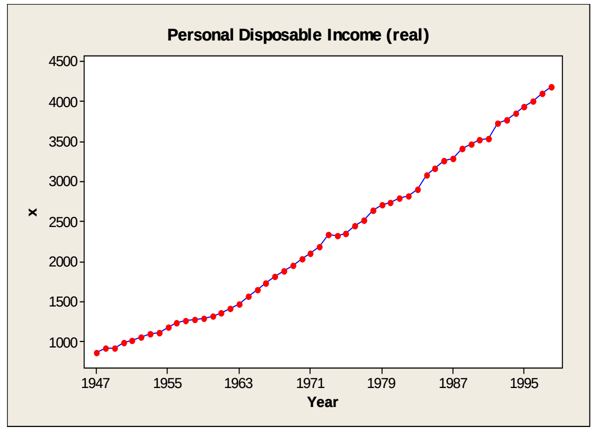

Fig. 14: t-plot of disposable income

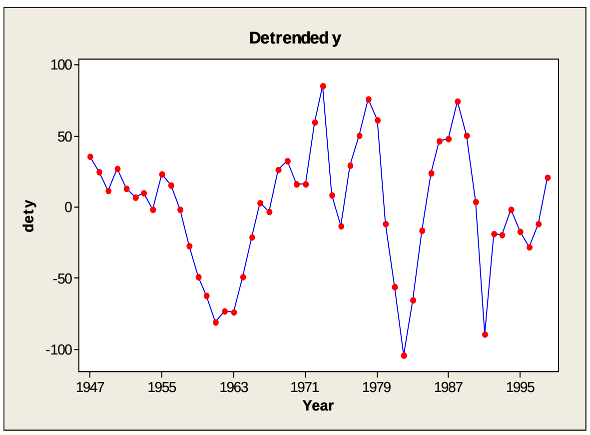

Fig. 15: Detrended consumers' expenditure

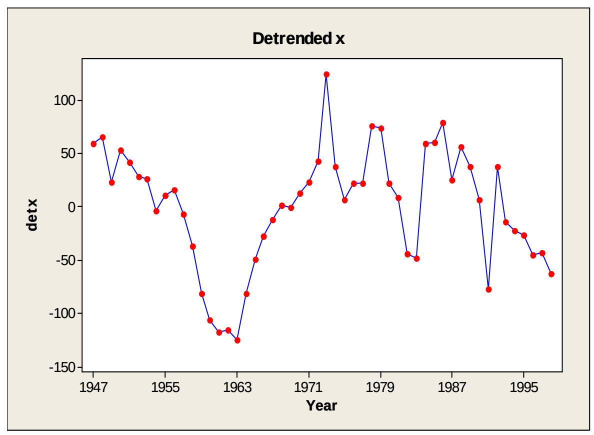

Fig. 16: Detrended disposable income

To assess the Normality assumption one needs to go a step further and subtract the temporal dependence by regressing the original data on both a trend polynomial and certain lags on (x_t, y_t), say (x_{t-i}, y_{t-i}), $i = 1, 2$. The de-trended and de-memorized (D/D) data are shown in figures 17-18, where one can detect a certain asymmetry and t-heterogeneity in the variation around the mean; the variance seems to increase with t. Moreover, the non-Normality (asymmetry) of the underlying joint distribution is also detectable in the scatter plot shown in figure 8. The detected departures from Normality are particularly crucial in the present case because, as shown in table 2, it might lead to drastic respecifications of the regression and skedastic functions.

In light of the above analogical reasoning, the best educated guess score-card for potentially appropriate reduction assumptions now reads:

(D) **Distribution**	(M) **Dependence**	(H) **Heterogeneity**
non-Normal	Markov	mean-heterogeneous
asymmetric		variance-heterogeneous?

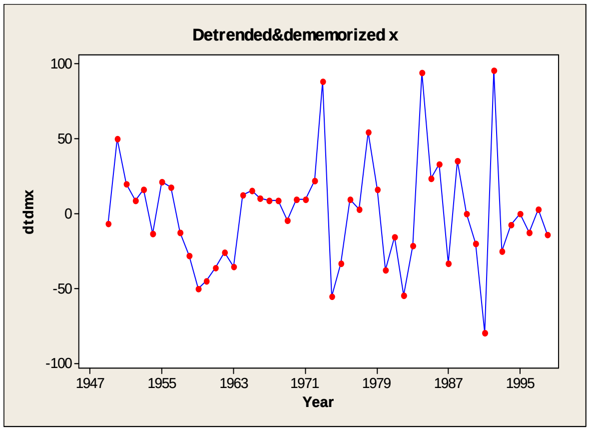

Fig. 17: D/D consumers' expenditure

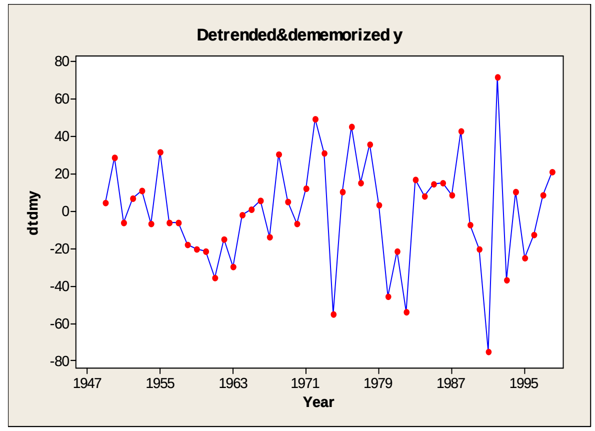

Fig. 18: D/D disposable income

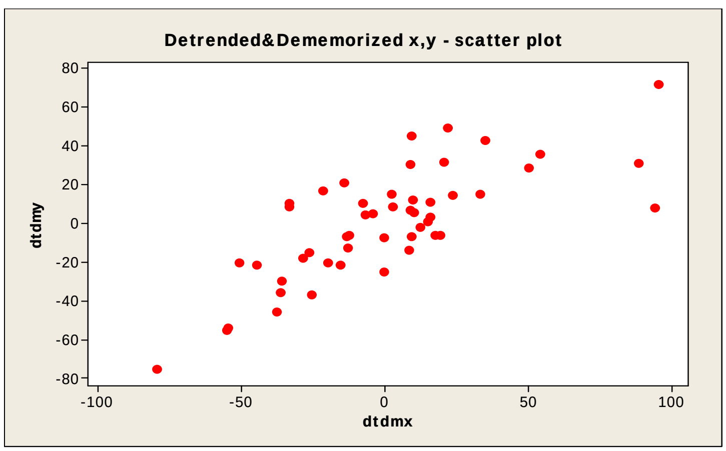

Fig. 19: Scatter-plot of detrended and demomorized series

In order to render the above informal assessment, relying on t-plots and scatter plots, more precise one can proceed to apply formal M-S tests. In particular, in an attempt to avoid 'erroneous' diagnoses as well as minimize the maintained assumptions, one can apply joint M-S tests specified in terms of auxiliary regressions based on the *studentized residuals:*

$$\widehat{v}_t = \frac{\sqrt{n}(y_k - \widehat{\beta}_0 - \widehat{\beta}_1 x_t)}{s},\ \ t{=}1,...,n.$$

The following system of auxiliary regressions is indicative of the type of M-S tests a PR modeler has in mind when probing for potential departures from model assumptions [1]-[5], denoted by a vertical line above; see Spanos (2010b).

$$
\left.\begin{array}{l}
\widehat{v}_t=\gamma_{10}+\gamma_{11}x_t+\overbrace{\gamma_{12}t+\gamma_{13}t^2}^{\overline{[5]}}+\overbrace{\gamma_{14}x_t^2}^{\overline{[2]}}+\overbrace{\gamma_{15}x_{t-1}+\gamma_{16}y_{t-1}}^{\overline{[4]}}+\varepsilon_{1t},\\
\widehat{v}_t^2=\gamma_{20}+\overbrace{\gamma_{21}x_t}^{\overline{[3]}}+\overbrace{\gamma_{22}t+\gamma_{23}t^2}^{\overline{[5]}}+\overbrace{\gamma_{24}x_t^2}^{\overline{[3]}}+\overbrace{\gamma_{25}x_{t-1}^2+\gamma_{26}y_{t-1}^2}^{\overline{[3]}-\overline{[4]}}+\varepsilon_{2t},\\
\widehat{v}_t^3=\overbrace{\gamma_{30}}^{\overline{[1]}}+\overbrace{\gamma_{31}x_t}^{\overline{[3]}}+\overbrace{\gamma_{32}t+\gamma_{33}t^2}^{\overline{[5]}}+\overbrace{\gamma_{34}x_t^2}^{\overline{[3]}}+\overbrace{\gamma_{35}x_{t-1}^2+\gamma_{36}y_{t-1}^2}^{\overline{[3]}-\overline{[4]}}+\varepsilon_{3t},\\
\widehat{v}_t^4=\overbrace{\gamma_{40}}^{\overline{[1]}}+\overbrace{\gamma_{41}x_t}^{\overline{[3]}}+\overbrace{\gamma_{42}t+\gamma_{43}t^2}^{\overline{[5]}}+\overbrace{\gamma_{44}x_t^2}^{\overline{[3]}}+\overbrace{\gamma_{45}x_{t-1}^2+\gamma_{46}y_{t-1}^2}^{\overline{[3]}-\overline{[4]}}+\varepsilon_{4t},
\end{array}\right\}\ t{=}1,2,...,n
$$

Note that most parametric M-S tests for Normality ([1]) rely on the skewness (α_3) and kurtosis (α_4) coefficients, like the D'Agostino-Pearson (1973) ($D'AP$) test in table 3, assume that the third and fourth central moments are constant, which might not be the case in practice. When these moments depend on any of the auxiliary terms indicated above, such Normality tests are likely to be misleading because the retained constancy of α_3 and α_4 is false . This is, indeed, the case for the estimated regression (16) above. Using a nonparametric M-S test for Normality ([1]), like the Anderson-Darling (A-D) test (Spanos, 1999) yields $A\text{-}D$=1.688[.005], which indicates departures from [1]. This brings out the importance of combining highly directional M-S tests (parametric), and omnibus (often nonparametric) to enhance the reliability of the diagnosis relying on M-S testing; see Spanos (2010b) for several enhancing M-S testing strategies.

Regression function. In view of the chance regularity patterns exhibited by the data in figures 1-8, the test that suggests itself would be based on the auxiliary regression:

$$
\widehat{v}_t=\gamma_0+\gamma_1x_t\quad\overbrace{+\gamma_2t}^{\text{trend}}\quad\overbrace{+\gamma_3x_t^2+}^{\text{non-linearity}}\quad\overbrace{\gamma_4x_{t-1}+\gamma_5x_{t-1}}^{\text{dependence}}\quad+v_t,
$$

and the following hypotheses of interest:

$$
H_0\text{: }\gamma_2{=}0,\ \gamma_3{=}0,\ \gamma_4{=}0,\ \gamma_5{=}0,\ \text{vs. }H_1\text{: }\gamma_2{\neq}0\text{ or }\gamma_3{\neq}0\text{ or }\gamma_4{\neq}0\text{ or }\gamma_5{\neq}0.
$$

The F test for these hypotheses yields $F(4,46)=50.682[.00000]^*$, where the contribution of each of the terms separately is evaluated to be:

Mean heterogeneity: $H_0:\gamma_2=0,\ F(1,46)=7.069[.011]^*$,

Non-linearity: $H_0:\gamma_3=0,\ F(1,46)=21.366[.000]^*$,

Temporal dependence: $H_0:\gamma_4=0,\ \gamma_5=0,\ F(2,46)=7.348[.002]^*$.

Skedastic function. The auxiliary regression that suggests itself is:

$$
\widehat{u}_t^2=\delta_0\quad\overbrace{+\delta_1t}^{\text{trend}}\quad\overbrace{+\delta_2x_t^2+\delta_3\widehat{u}_{t-1}^2}^{\text{heteroskedasticity}}\quad+v_t,
$$

The F test for the joint significance of the terms t, x_t^2 and $\widehat{u}_{t-1}^2$ yields $F(4,46)$=27.630[.000]*, where the contribution of each of the terms separately is:

Variance heterogeneity: H_0: $\delta_1 = 0,\ F(1,46) = 3.201[.091]$,

Heteroskedasticity: H_0: $\delta_2{=}0,\ \delta_3{=}0,\ F(2,46){=}11.901[.000]^*$.

These two M-S test results imply that the misspecification present takes the form of heteroskedasticity, and not heterogeneity! To bring out the importance of joint M-S tests, consider the case where the auxiliary regression does not include a heteroskedasticity term:

$$\widehat{u}_t^2 = \delta_0 + \delta_1 t + v_t.$$

The M-S test result for the significance of δ_1 would have yielded $F(1,46){=}9.085[.004]^*$, which leads to the erroneous diagnosis that the conditional variance is trending.

5.3 Respecification that ignores the non-Normality

To bring out the importance of taking the distributional assumption seriously, consider a scenario where a modeler ignored the departures from *Normality* indicated in figures 17-19, but respecified to account for the trends in mean and temporal dependence. The respecified model called for in such a scenario is the Dynamic Linear Regression (DLR) model with a trend given in table 4.

Table 4 - Normal/Dynamic Linear Regression (N/DLR) Model

Statistical GM:		$y_t = \beta_0 + \delta t + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1} + u_t, \quad t\in\mathbb{N}.$	
[1]	Normality:	$(y_t\vert x_t, x_{t-1}, y_{t-1}) \backsim \mathsf{N}(.,.)$,	$t\in\mathbb{N}.$
[2]	Linearity:	$E(y_t\vert x_t, x_{t-1}, y_{t-1}) = \beta_0 + \delta t + \beta_1 x_t + \beta_2 y_{t-1} + \beta_3 x_{t-1}$,	
[3]	Homo/sticity:	$Var(y_t\vert x_t, x_{t-1}, y_{t-1}) = \sigma_0^2$,	
[4]	Markov:	$\{(y_t\vert x_t),\ t\in\mathbb{N}\}$ Markov process,	
[5]	t-invariance:	$(\beta_0, \beta_1, \beta_2, \beta_3, \sigma_0^2)$ are t-constant,	

When the N/DLR model is estimated it yields:

$$y_t = \underset{(50.29)}{109.91} + \underset{(2.289)}{1.785}t + \underset{(.049)}{.129}t^2 + \underset{(.078)}{.651}x_t + \underset{(.119)}{.724}y_{t-1} - \underset{(.109)}{.513}x_{t-1} + \widehat{\varepsilon}_t,$$
$$R^2 = .9997,\ s = 20.076,\ T = 52. \tag{19}$$

The end result of such a scenario is that (19) is likely to be misspecified in particular ways. In particular, a PR modeler could use table 2 to anticipate that any departures from Normality are likely to give rise to departures from the *Linearity* ([2]) and *Homoskedasticity* ([3]) assumptions in table 3. Indeed, this is confirmed by the M-S testing results shown in table 5.

Table 5 - Mis-Specification (M-S) tests for (19)	
Non-Normality:	$D'AP = .347[.841]$
Non-linearity:	$F(1,45) = 14.933[.000]^*$
Heteroskedasticity:	$F(2,43) = 7.137[.002]^*$
Autocorrelation:	$F(1,44) = 1.438[.237]$
t-heterogeneity:	$F(1,45) = 0.781[.476]$

It is interesting to note that the autocorrelation-corrected model (M_1) in (17) is a highly resticted form of the above N/DLR model; see McGuirk and Spanos (2009). Hence, the M-S testing results in table 5 are also relevant for M_1.

5.4 Respecification: the Probabilistic Reduction approach

How would a PR modeler proceed in light of the M-S testing results that all five assumptions [1]-[5] of the N/LR model (table 1) are invalid for annual time USA data $\mathbf{z}_0$? In a sense, respecification requires one to go back to the drawing board and replace the original reduction assumptions with ones that account for the regularities observed in figures 13-19, as well as the overall picture painted by the various M-S testing results. The one key result from the latter is that the departure from the assumed conditional variance was not t-heterogeneity but heteroskedasticity. This provides a clue to the non-Normality required to account for the detected departure. It is well-known that the logarithmic transformation can often be used to stabilize the conditional variance (see Spanos, 1986, pp. 487-9). This suggests that one might consider taking logs of the original variables to respecify the original model. To assess such a respecification one could have another look at the various t-plots and scatter plots of the transformed data $\ln \mathbf{z}_0$. Figures 20-21 show the t-plots of $\ln \mathbf{z}_0$. Of particular interest, however, are the t-plots of de-trended and de-memorized (D/D) data $\ln \mathbf{z}_0$, as well as their scatter plot shown in figures 22-24.

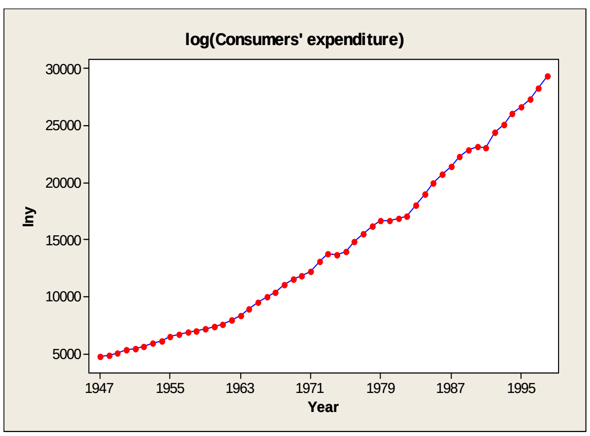

Fig. 20: t-plot of $\ln y_t$

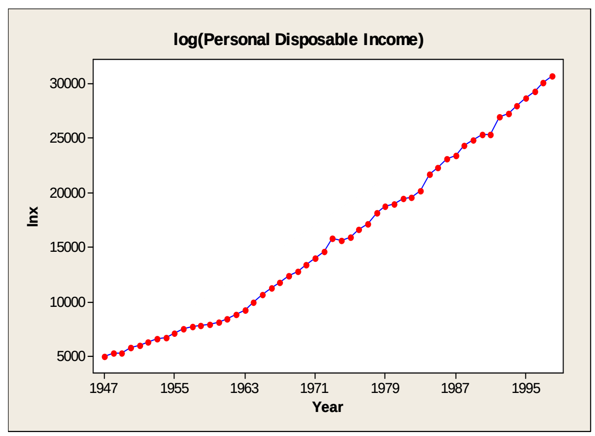

Fig. 21: t-plot of $\ln x_t$

The t-plots of D/D $\ln \mathbf{z}_0$ do not exhibit any of the original t-heterogeneity in the variance as in figures 17-18, and the scatter plot in figure 24 exhibits no major departures from the elliptically shaped symmetry associated the bivariate Normal distribution.

The above graphical assessment indicates that a sensible respecification of the original N/LR model to account for the systematic statistical information ignored by that model, might rely on the Reduction assumptions:

(D) **Distribution**	(M) **Dependence**	(H) **Heterogeneity**
Normal (in logs)	Markov	mean-heterogeneous

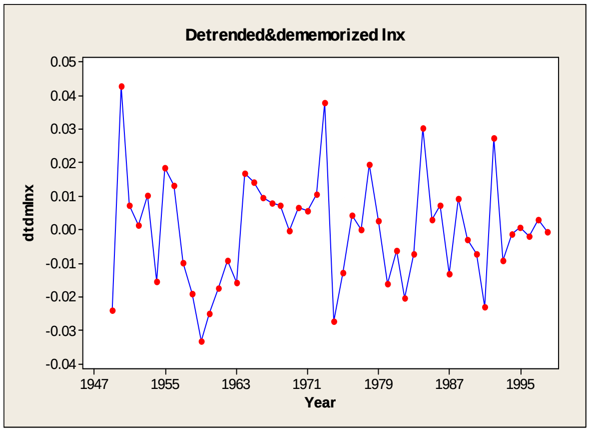

Fig. 22: t-plot of D/D $\ln y_t$

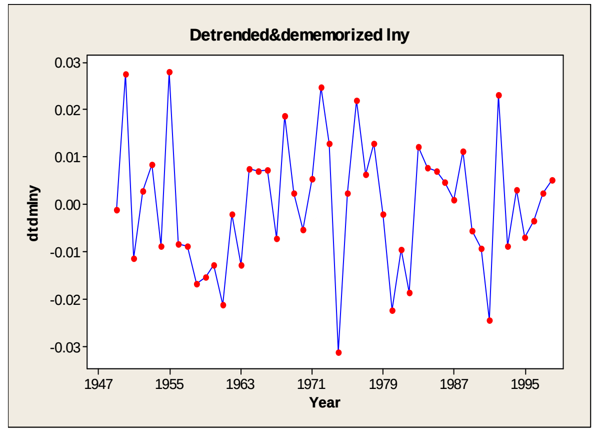

Fig. 23: t-plot of D/D $\ln x_t$

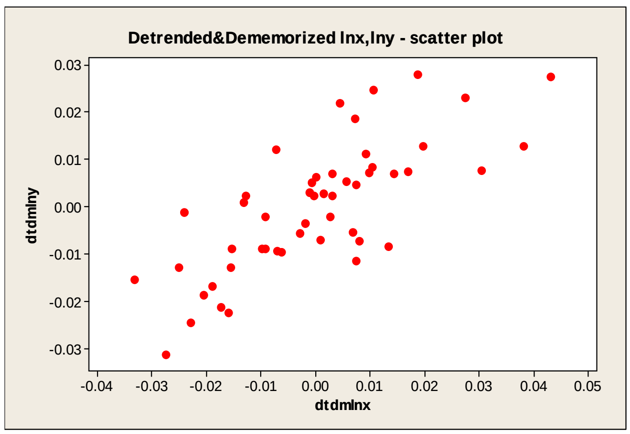

Fig. 24: scatter plot of D/D $\ln y_t$ and $\ln x_t$

Such a respecification will give rise to the DLR model with a trend (table 4) based on $\ln \mathbf{Z}$:=$(\ln x_t, \ln y_t)$. Estimating this statistical model yields:

$$\ln y_t = \underset{(.272)}{.912} + \underset{(.001)}{.005t} + \underset{(.069)}{.708} \ln x_t + \underset{(.108)}{.565} \ln y_{t-1} - \underset{(.097)}{.413} \ln x_{t-1} + \widehat{\varepsilon}_t, \tag{20}$$

$$R^2{=}.9997, \ s = .0084, \ T = 52.$$

A subset of M-S test results reported in table 6 indicate no significant departures from the assumptions of this model (see Spanos, 1986, ch. 23).

Table 6 - Mis-Specification (M-S) tests for (20)	
Non-Normality:	$D'AP = .872[.646]$
Non-linearity:	$F(1,45) = 1.142[.291]$
Heteroskedasticity:	$F(2,43) = .317[.730]$
Autocorrelation:	$F(1,44) = 1.129[.293]$
t-heterogeneity:	$F(2,43) = .423[.671]$

The statistically adequate model (20) provides no evidence *for* the AIH. Indeed, it calls into question both the choice of variables as well as the missing dynamics. The question, however, is: does (20) also constitute a substantively adequate model?

5.5 What about substantive adequacy?

On the basis of a thorough M-S testing one can claim that the statistically adequate model (20) accounts for the statistical regularities in the data, but it does *not* constitute a substantive model as it stands. The presence of the trend indicates *substantive ignorance* by suggesting that certain relevant (substantive) variables are clearly missing. The use of a trend (t) in a regression-type model provides a generic way to 'capture' t-heterogeneity, but it does nothing to *explain* why the data generating mechanism gives rise to data that exhibit such regularities. Returning to macro-economic theory one can pose the question of omitted relevant variables in the context of the statistically adequate model (20) and use the error reliability of the model to probe the significance of potential variables; see Spanos (2006b). In this particular case, the variables selected to account for the trend are: $x_{2t}-$ price level, $x_{3t}-$ consumer credit outstanding, $x_{4t}-$ short run interest rate; see Economic Report of the President (2000).

The details of how reliable statistical procedures, in the context of a statistically adequate model, can guide one to an empirical model are beyond the scope of this paper, but see Spanos (2006b). Using such a strategy one can proceed from (20) to select an equilibrium-correction empirical model (see Hendry, 2000) that accounts for the regularities in the data (statistically adequate) and enjoys a clear substantive interpretation:

$$\begin{array}{l}\Delta \ln y_t = \underset{(.004)}{.004} - \underset{(.046)}{.199}(\ln y_t - \ln x_{1t}) + \underset{(.056)}{.558}\Delta \ln x_{1t} - \underset{(.071)}{.173}\Delta \ln x_{2t} + \\ +\underset{(.025)}{.081}\Delta \ln x_{3t} - \underset{(.011)}{.024}\Delta \ln x_{4t} + \widehat{\varepsilon}_t,\ R^2 = .924,\ s = .0071,\ T = 52.\end{array} \tag{21}$$

One might dispute the choice and substantive relevance of the particular additional variables chosen above, but there is nothing wrong with that. Indeed, such challenges should constitute a crucial component of an on-going dialogue between econometricians and economic theorists that needs to be established for economics to become a real scientific discipline that aspires to have solid empirical foundations.

6 Summary and conclusions

The paper focused primarily on bringing out a crucial methodological problem that has bedeviled empirical research with observational data since the early 20th century, and proposing particular ways to address the problem at the level of a practitioner. In a nutshell, the problem is that when one's favorite theory is foisted on the data, the end result is invariably an empirical model which is both statistically and substantively misspecified, but one has no way to disentangle the two sources of error. How is a modeler to do when faced with such a scenario?

Ignoring the problem of statistical misspecification, and focusing exclusively on the evaluation of the statistical results — taken at face value — on substantive grounds, has proved a disastrous strategy for learning from data. Essentially, one is using statistically unreliable inference procedures [the actual error probabilities are very different from the nominal ones] to forge untrusworthy evidence that have no bearing and shed no light on the phenomenon of interest. The results of such a strategy are calamitous enough to be described as the hush scandal of applied econometrics of the last century! Instead of genuine 'learning from data' applied econometrics has become a story-telling enterprise relying on meaningless numbers, masquerading as reliable inference results, forged using computer packages.

It is argued that the only sensible way forward is to address the statistical misspecification problem before any substantive questions of interest can be reliably appraised. However, separating the statistical and substantive information in empirical modeling with observational data has been a major obstacle to real progress for almost a century. The key to circumventing this problem is to come up with a purely probabilistic construal of the notion of a statistical model $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ whose specification relies solely on the statistical information contained in the data, but allows for the embedding of the substantive model in its context; see Spanos (1986). Viewing $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ as a parametrization of probabilistic structure of observable stochastic process $\{\mathbf{Z}_t,\ t\in\mathbb{N}\}$ underlying the data $\mathbf{z}_0$, enables one to secure its statistical adequacy vis-a-vis $\mathbf{z}_0$ using thorough Mis-Specification (M-S) testing guided by discerning graphical techniques and effective probing strategies. The substantive and statistical information are reconciled by embedding the structural model $\mathcal{M}_{\boldsymbol{\varphi}}(\mathbf{x})$ in the context of a statistically adequate $\mathcal{M}_{\boldsymbol{\theta}}(\mathbf{z})$ and assessing the restrictions $\mathbf{G}(\boldsymbol{\theta},\boldsymbol{\varphi})=\mathbf{0}$.

This calls for drastic changes to the traditional textbook methodological framework with a view to separate the statistical from the substantive information, at least ab initio, and secure statistical adequacy before any appraising of substantive adequacy. These modeling strategies are in direct conflict with the traditional textbook ad hoc *error-fixing* practices, such as error-autocorrelation and heteroskedasticity corrections or trying out different estimation methods (OLS, GLS, IV, etc.), that do little, if anything, to address the unreliability of inference problem; see Spanos and McGuirk (2001). Equally ineffective are modeling strategies that rely on weakening the statistical premises and invoking asymptotic inference, as well as nonparametric methods that replace testable with non-testable assumptions.

References

[1] Billingsley, P. (1995), *Probability and Measure*, 4th ed., Wiley, NY.

[2] Cox, D. R. (1990), "Role of Models in Statistical Analysis," *Statistical Science*, **5**: 169-174.

[3] Cox, D. R. and D. V. Hinkley (1974), *Theoretical Statistics*, Chapman & Hall, London.

[4] D'Agostino, R. B. and E. S. Pearson (1973), "Tests for departure from Normality: empirical results for the distribution of b_2 and $\sqrt{b_3}$," *Biometrika*, **60**: 613-622.

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

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

[7] Fisher, R. A. (1925), "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, **22**: 700-725.

[8] Fisher, R. A. (1934), "Two New Properties of Maximum Likelihood," *Proceedings of the Royal Statistical Society*, A, **144**: 285-307.

[9] Fisher, R. A. (1955), "Statistical methods and scientific induction," *Journal of the Royal Statistical Society*, **B**, **17**: 69-78.

[10] Friedman, M. (1953), "The Methodology of Positive Economics," pp. 3-43 in *Essays in Positive Economics*, Chicago University Press, Chicago; reprinted in Mäki (2009).

[11] Goldberger, A.S. (1964), *Econometric Theory*, Wiley, NY.

[12] Guala, F. (2005), *The Methodology of Experimental Economics*, Cambridge University Press, Cambridge.

[13] Gujarati, D. N. (2009), *Basic Econometrics*, 5th ed., McGraw-Hill, NY.

[14] Greene, W. H. (2008), *Econometric Analysis*, 6th ed., Prentice Hall, NJ.

[15] Hacking, I. (1965), *Logic of Statistical Inference*, Cambridge University Press, Cambridge.

[16] Hendry, D. F. (2000), *Econometrics: Alchemy or Science?*, 2nd ed., Blackwell, Oxford.

[17] Johnston, J. (1963), Econometric Methods, McGraw-Hill, NY.

[18] Kennedy, P. (2008), *A Guide to Econometrics*, 6th edition, MIT Press, MA.

[19] Keynes, J. M. (1939), "Professor Tinbergen's Method," *Economic Journal*, **49**: 558-568.

[20] Koutris, A., M. Heracleous and A. Spanos (2008), "Testing for Non-stationarity using Maximum Entropy Resampling: a Misspecification Testing Perspective," *Econometric Reviews*, **27**: 363-384.

[21] Kramer, W. and H. Sonnberger (1986), *The Linear Regression Model under Test*, Physica-Verlag Heidelberg Wien.

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

[23] Li, Q. and J. S. Racine (2006), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, NJ.

[24] Mäki, U. (2009), *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, Cambridge University Press, Cambridge.

[25] Malthus, T. R. (1936), Principles of Political Economy, Reprints of Economic Classics, 1986, Augustus, M. Kelley, NJ.

[26] Matyas, L. (ed.) (1999), *Generalized Method of Moments Estimation*, Cambridge University Press, Cambridge.

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

[28] Mayo, D. G. (1997), "Duhem's Problem, the Bayesian Way, and Error Statistics, or "What's Belief Got to Do with It?", *Philosophy of Science*, **64**, 222-244.

[29] Mayo, D. G. and A. Spanos (2004), "Methodology in Practice: Statistical Misspecification Testing", *Philosophy of Science*, **71**: 1007-1025.

[30] Mayo, D. G. and A. Spanos. (2006), "Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction," *The British Journal for the Philosophy of Science,* **57**: 323-357.

[31] Mayo, D. G. and Spanos, A. (eds.) (2010a), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability and the Objectivity and Rationality of Science*, Cambridge University Press, Cambridge.

[32] Mayo, D. G. and A. Spanos (2010b), "Error Statistics," forthcoming in *Philosophy of Statistics*, the *Handbook of Philosophy of Science*, Elsevier (editors) D. Gabbay, P. Thagard, and J. Woods.

[33] McGuirk, A. and A. Spanos (2009), "Revisiting Error Autocorrelation Correction: Common Factor Restrictions and Granger Non-Causality," *Oxford Bulletin of Economics and Statistics,* **71**: 273-294.

[34] Mills, T.C. and K. Patterson, (2006), *New Palgrave Handbook of Econometrics*, vol. 1, MacMillan, London.

[35] Moore, H. L. (1914), *Economic Cycles - Their Laws and Cause*, McMillan, NY.

[36] Neyman, J. (1937), "Outline of a Theory of Statistical Estimation based on the Classical Theory of Probability," *Philosophical Transactions of the Royal Statistical Society of London*, A, **236**: 333–380.

[37] Neyman, J. (1956), "Note on an Article by Sir Ronald Fisher," *Journal of the Royal Statistical Society*, B, **18**: 288-294.

[38] 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.

[39] Pagan, A.R. (1987), "Three econometric methodologies: a critical appraisal", *Journal of Economic Surveys,* **1**, 3-24. Reprinted in C. W. J. Granger (1990).

[40] Pagan, A.R. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

[41] Pearson, K. (1920), "The Fundamental Problem of Practical Statistics," *Biometrika*, **XIII**, 1-16.

[42] Pearson, E. S. (1955), "Statistical Concepts in the Relation to Reality," J*ournal of the Royal Statistical Society, Series B*, **17**, 204-207.

[43] Rao, C. R. (2004), "Statistics: Reflections on the Past and Visions for the Future," *Amstat News*, **327**: 2-3.

[44] Ricardo, D. (1817), Principles of Political Economy and Taxation, vol. 1 of The Collected Works of Davie Ricardo, ed. P. Sraffa and M. Dobb, Cambridge University Press, Cambridge.

[45] Spanos, A., (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.

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

[47] Spanos, A. (1990), "The Simultaneous Equations Model revisited: statistical adequacy and identification", *Journal of Econometrics*, **44**: 87-108.

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

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

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

[51] Spanos, A. (2001), "Parametric versus Non-parametric Inference: Statistical Models and Simplicity," ch. 11, pp. 181-206 in *Simplicity, Inference and Modelling: Keeping it Sophisticatedly Simple*, edited by A. Zellner, H. A. Keuzenkamp and M. McAleer, Cambridge University Press.

[52] Spanos, A. (2006a), "Econometrics in Retrospect and Prospect," pp. 3-58 in Mills, T.C. and K. Patterson, *New Palgrave Handbook of Econometrics*, vol. 1, MacMillan, London.

[53] Spanos, A. (2006b), "Where Do Statistical Models Come From? Revisiting the Problem of Specification," pp. 98-119 in *Optimality: The Second Erich L. Lehmann Symposium*, edited by J. Rojo, Lecture Notes-Monograph Series, vol. 49, Institute of Mathematical Statistics.

[54] Spanos, A. (2006c), "Revisiting the omitted variables argument: substantive vs. statistical adequacy," *Journal of Economic Methodology*, **13**: 179–218.

[55] Spanos, A. (2007a), "The Instrumental Variables Method revisited: On the Nature and Choice of Optimal Instruments," pp. 34-59 in *Refinement of Econometric Estimation and Test Procedures*, ed. by G. D. A. Phillips and E. Tzavalis, Cambridge University Press, Cambridge.

[56] Spanos, A. (2007b), "Curve-Fitting, the Reliability of Inductive Inference and the Error-Statistical Approach," *Philosophy of Science*, **74**: 1046–1066.

[57] Spanos, A. (2008), "Statistics and Economics," pp. 1129-1162 in the *New Palgrave Dictionary of Economics*, 2nd edition, Eds. S. N. Durlauf and L. E. Blume. Palgrave Macmillan, London.

[58] Spanos, A. (2009a), "The Pre-Eminence of Theory versus the European CVAR Perspective in Macroeconometric Modeling," *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 3, 2009-10. http://www.economics-ejournal.org/economics/journalarticles/2009-10.

[59] Spanos, A. (2009b), "Statistical Misspecification and the Reliability of Inference: the simple t-test in the presence of Markov dependence," forthcoming in *Korean Economic Review.*

[60] Spanos, A. (2010a), "Theory Testing in Economics and the Error Statistical Perspective," pp. 202-246 in *Error and Inference*, edited by Mayo, D. G. and A. Spanos (2010a).

[61] Spanos, A. (2010b), "Akaike-type Criteria and the Reliability of Inference: Model Selection vs. Statistical Model Specification," forthcoming, *Journal of Econometrics.*

[62] Spanos, A. and A. McGuirk (2001), "The Model Specification Problem from a Probabilistic Reduction Perspective," *Journal of the American Agricultural Association*, **83**: 1168-1176.

[63] Spanos, A. and A. McGuirk (2002), "The Problem of Near-Multicollinearity Revisited: Erratic vs. Systematic Volatility", *Journal of Econometrics*, **108**: 365-393.

[64] Stock, J. H. and M. W. Watson (2006), *Introduction to Econometrics*, 2nd ed., Addison-Wesley.

[65] Swamy, P. A. V. B. (1970), "Efficient inference in a random coefficient regression model," *Econometrica*, **38**: 311-323.

[66] Swamy, P. A. V. B. (1971), *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, NY.

[67] Wooldridge, J. M. (2009), *Introductory Econometrics: a Modern Approach*, 3rd ed., Thomson, South-Western.