Revisiting Haavelmo’s Structural Econometrics: Bridging the Gap between Theory and Data

Aris Spanos
Department of Economics,
Virginia Tech, Blacksburg, VA 24061

December 2011 - First draft

Abstract

The primary aim of this paper is threefold. First, to explain the neglect of Haavelmo’s methodological ideas and insights in current textbook econometrics as largely symptomatic of the fact that they are at odds with the current empirical modeling framework in economics. Second, to argue that this neglect has contributed significantly to rendering the overwhelming majority of published empirical evidence untrustworthy, and thus derailing any learning from data about economic phenomena of interest and undermining the credibility of the empirical foundations of economics. Third, to make a case that by blending some of Haavelmo’s key methodological ideas and insights into a refined/extended Fisher-Neyman-Pearson statistical approach to inference, gives rise to a more pertinent modeling framework for observational data. The key to this blending is the distinction between substantive and statistical information, which is used to shed light on several confusions permeating textbook econometrics. This gives rise to a framework that can ensure learning from data by securing the reliability of inference by validating the underlying statistical model.
1 Introduction

Haavelmo was awarded the Nobel Prize in Economics in 1989 “for his clarification of the probability theory foundations of econometrics and his analyses of simultaneous economic structures”. A glance through the current traditional textbooks in econometrics reveals that most of them do not even mention Haavelmo [Davidson and MacKinnon, 2004; Rudd, 2000; Stock and Watson, 2011; Wooldridge, 2009, inter alia], and the few that do [Greene, 2011; Wooldridge, 2010] only cite Haavelmo (1943). It is interesting to note that none of them credit him for contributing to ‘clarifying the probability theory foundations of econometrics’. In this sense the current textbook literature gives the impression that Haavelmo’s only contribution to current econometrics was to solve the technical problem of least-squares bias arising when estimating simultaneous equations. Therefore, to the extent that these textbook articulate the current conventional wisdom, his numerous methodological ideas and insights pertaining to econometric modeling (especially Haavelmo, 1944), are considered largely irrelevant to the current practice. This is despite the early highly positive reaction by the Cowles Commision group [Koopmans, 1950; Hood and Koopmans, 1953; Epstein, 1987; Hendry and Morgan (1995); Morgan 1990; Qin, 1993], as well as the more recent literature highlighting Haavelmo’s multifaceted contributions [Aldrich, 1989, 1994; Anderson, 1991; Bjerkholt, 2005, 2007; Heckman, 1992; Hendry et al, 1989; Moene and Rodseth, 1991; Nerlove, 1990; Spanos, 1986, 1988, 1989]. The role played by the Cowles Commission in distilling Haavelmo’s ideas and influencing the early textbook perspective (Johnston, 1963; Goldberger, 1964) is beyond the scope of the present paper, but see Spanos (2010a).

The primary aim of this paper is threefold.

First, to explain the neglect of Haavelmo’s methodological ideas and insights as largely symptomatic of the fact that they are at odds with the current approach to empirical modeling in economics.

Second, to argue that this neglect has contributed significantly to rendering the overwhelming majority of published empirical evidence untrustworthy, derailing any learning from data about economic phenomena of interest as well as undermining the credibility of any empirical foundations of economics.

Third, to make a case that by blending some of Haavelmo’s key methodological ideas and insights into a refined/extended Fisher-Neyman-Pearson (F-N-P) statistical approach to inference, one can construct a more pertinent modeling framework for empirical modeling in economics. A framework that can ensure learning from data by securing the reliability of inference.

The discussion that follows will inevitably be idiosyncratic because it relates to how Haavelmo’s ideas and insights have influenced the development of my own thinking about empirical modeling in economics since my graduate students days at the London School of Economics, where I came across Haavelmo (1944) accidentally!

Sections 2 and bring out the fact that Haavelmo’s probabilistic perspective on statistical models, relating the data to the joint distribution of the observables —
2 Haavelmo’s ‘Probability Approach’ and the textbook perspective

Historically, theory has generally held the pre-eminent role in empirical modeling in economics with data being given the subordinate role of facilitating ‘the quantification of theories’ presumed to be true; see Spanos (2010a). Typically, a traditional textbook modeler begins with a deterministic theory model, and transforms that into a statistical (econometric) model by:

[i] viewing the theory model as the systematic component and
[ii] attaching probabilistic error terms to define the non-systematic component.

This perspective preserves the pre-eminence of theory in the form of deterministic models, and offers a way to use statistical inference techniques by attaching ‘random’ error (shock) terms to specify structural models.

This is exactly the approach Haavelmo (1944) was calling into question:

"So far, the common procedure has been, first to construct an economic theory involving exact functional relationships, then compare this theory with some actual measurements, and, finally, "to judge" whether the correspondence is "good" or "bad". Tools of statistical inference have been introduced, in some degree, to support such judgements, e.g. the calculation of a few standard errors and multiple-correlation coefficients. The application of such simple "statistics" has been considered legitimate, while, at the same time, the adoption of definite probability models has been deemed a crime in economic research, a violation of the very nature of economic data.” (p. iii)

The prevailing view in economics at that time was that statistical techniques could only be applied to data that could be realistically viewed as realizations of "random samples", i.e. Independent and Identically Distributed (IID) random variables; see Frisch (1934), p. 6. However, the systematic development of the theory of stochastic
processes in the 1930s, due primarily to Kolmogorov and Kintchin (see Doob, 1953), significantly extended the intended scope of statistical modeling; see Mann and Wald (1943). Haavelmo knew first hand about the new developments in frequentist statistics because he spent enough time with Jerzy Neyman in 1936 in London and Berkeley in 1939, as well as Abraham Wald during his stay (1939-1944) in the USA; see Bjerkholt (2007). Hence, it is not surprising that he was the first to articulate these developments to econometricians:

"... it is not necessary that the observations should be independent and that they should follow the same one-dimensional probability law. It is sufficient to assume that the whole set of, say $n$, observations may be considered as one observation of $n$ variables (or a "sample point") following an $n$-dimensional joint probability law, the "existence" of which my be purely hypothetical. ... Modern statistical theory has made considerable progress in solving such problems of statistical inference." (Haavelmo, 1944, pp. iii-iv)

By the 1960s most economists understood that statistical techniques can be applied to non-IID data, like economics time series, but the broader probabilistic perspective of viewing the data as a realization of a stochastic process, stemming from the Fisher-Neyman-Pearson (F-N-P) perspective, was never integrated into traditional textbook econometrics as the presumptions [i]-[ii] above attest. Why did Haavelmo reject the presumptions [i]-[ii]? In his Nobel lecture Haavelmo gives full credit to Jerzy Neyman for his conversion from the Frisch errors-in-variables scheme to the F-N-P perspective; Haavelmo (1989).

The current textbook perspective contravenes both the Haavelmo ‘Probability Approach’ viewpoint and the F-N-P perspective, which consider the stochastic nature of data to be intrinsic. That is, probability is attached directly to the observable random variables underlying the data and viewing the latter as a single realization of the sample whose probabilistic structure is specified by their joint distribution, which provides the basis of all inductive inferences; hence the key role of the likelihood function. Indeed, Haavelmo at the 1942 Hillside Econometric Seminar presentation included this at the top of the list of several interesting methodological questions:

"1. Is there any danger in considering economic data as stochastic variables? Is there any use in so doing?
   2. Is it always objectionable to use freehand methods of curve fitting?
   3. (a) What is the meaning of a “spurious result”?
      (b) What are “economically meaningful” results, as contrasted with “statistically significant”?
   4. Is it more dangerous, in general, to draw inference from economic time series than from other kinds of statistical information?” (see Haavelmo and Bjerkholt, 2007, p. 840).

Parenthetically, it is important to note that the above quotation raises several crucial methodological issues most of which linger on to this day. Most notable are the questions pertaining to ‘spurious results’ and the distinction between ‘substantive (economically meaningful) and statistical significance’.

In light of the above discussion, one can make a strong case that by focusing
on the ‘quantification of theoretical relationships’ textbook econometrics reverted to the pre-Fisher curve-fitting framework where the probabilistic structure enters the modeling in a non-intrinsic way thorough non-systematic terms that represent errors of measurement, errors of approximation and ‘accidental’ omitted effects. In his most cited paper, Haavelmo (1943) warned econometricians about the perils of probabilistic modeling via error terms:

"... the notion that one can operate with some vague idea about "small errors" without introducing the concepts of stochastical variables and probability distributions, is, I think, based upon on illusion." (Haavelmo, 1943, p. 5)

In early 20th century statistics amounted to a toolkit of descriptive techniques and curve-fitting methods using least-squares for summarizing economic data. These descriptive methods, associated with the works of Graunt, Petty, Quetelet, Galton and Karl Pearson, were largely disjoined from the mathematical theory of probability; see Stigler (1986). To confuse matters, however, their results were often accompanied by ill-defined allusions to inductive inferences using probabilistic terminology like ‘mean’, ‘variance’, ‘standard deviation’, ‘correlation’ and ‘regression’, but invariably conflating the probabilistic concepts with their descriptive analogues in terms of the data; see Fisher (1922), p. 310-311. Proper integration of probability theory with statistical inference was initiated in the 1920s by R. A. Fisher.

3 The Fisher-Neyman-Pearson probabilistic perspective

Fisher (1922) pioneered modern frequentist statistics as a model-based approach to statistical induction anchored on the notion of a statistical model:

\[
\mathcal{M}_\theta(z) = \{ f(z; \theta), \ \theta \in \Theta \}, \ z \in \mathbb{R}^n, \ \Theta \subset \mathbb{R}^m, \ m < n, \tag{1}
\]

where \( f(z; \theta) \) is the distribution of the sample \( Z := (Z_1, ..., Z_n) \). He proposed to begin with a prespecified \( \mathcal{M}_\theta(z) \) (a ‘hypothetical infinite population’), and view data \( Z_0 := (z_1, ..., z_n) \) as a realization thereof. He viewed the specification of \( \mathcal{M}_\theta(z) \) as a response to the question:

"Of what population is this a random sample?" (ibid., p. 313), underscoring that:

"the adequacy of our choice may be tested a posteriori." (p. 314)

The most enduring contribution of the F-N-P perspective comes in the form of rendering the errors for statistical induction ascertainable by embedding the material experiment into a statistical model in terms of which the relevant error probabilities are defined, thus providing a measure of the reliability of the inference procedure; see Spanos (2006a). Learning from data is achieved when an inference is reached by an inductive procedure which, with high probability, will yield true conclusions from valid inductive premises (a statistical model).

It should be noted that the transition from the early 20th century descriptive statistics, relying on vague large sample (as \( n \to \infty \)) justifications, to statistical inference proper, in the hands of Fisher (1922, 1925, 1935) and Neyman and Pearson
(1933), relying on a given \( n > 1 \), was neither apparent nor without its polemists. It’s no accident that several statistics textbooks that were influential in economics, like Bowley (1937), Mills (1938) and Allen (1949), missed the change of paradigms.

Haavelmo (1943, 1944) was the first to attempt to integrate the F-N-P perspective into econometrics by:

[A] advocating a probabilistic foundation for inference based on the \( n \)-dimensional distribution \( D(Z_1, Z_2, ..., Z_n; \theta) \) of the observable (vector) process \( \{Z_t, t \in \mathbb{N} : (1, 2, ..., n, ...)\} \) underlying data \( Z_0 := (z_1, z_2, ..., z_n) \):

"When we have observed a set of values of \( n \) observable variables \( (z_1, z_2, ..., z_n) \) we may, without any possibility of a contradiction, say that these \( n \) values represent a sample point drawn from a universe obeying some unknown \( n \)-dimensional (integral) probability law." (Haavelmo, 1944, p. 48)

"... to avoid inconsistencies ... all formulae for estimating the parameters involved should be derived on the basis of this joint probability law of all the observable variables involved in the system." Pointedly adding: "(This, I think, is obvious to statisticians, but it is overlooked in the work of most economists who construct dynamic models to be fitted to economic data.)" (Haavelmo, 1943, p. 7)

In this sense, the probabilistic structure providing the inductive premises of inference, as envisaged by both Fisher and Haavelmo, is primarily intrinsic to the data \( Z_0 \), and thus viewed as a realization of a generic stochastic process \( \{Z_t, t \in \mathbb{N}\} \). This is in direct contrast to the current textbook econometric perspective where the probabilistic structure is theory-driven because it is introduced into an essentially deterministic theory model via latent (unobservable) error and shock terms. An obvious point that is often neglected by advocates of the textbook approach, is that the probabilistic premises that matter for the reliability of the inference procedures is not that of the error (shock) terms, but what that implies for the observable process \( \{Z_t, t \in \mathbb{N}\} \), i.e. the probabilistic assumptions pertaining to the observables that constitute the statistical model \( M_\theta(z) \):

"... all comes down to one and the same thing, namely, to study the properties of the joint probability distribution of the random (observable) variables in a stochastic equations system." (p. 85)

Hence, when the latter is imposed indirectly, via the error terms, and rarely validated against the data, it will invariably give rise to unreliable inferences and untrustworthy evidence. In practical terms this implies that the nominal (assumed) error probabilities used in any inferences made on the basis of the estimated model are likely to be very different from the actual error probabilities. The surest way to lead an inference astray is to use a 5% significance level test when the actual type I error probability is closer to 90%.

As argued below, the textbook perspective is at odds not just with Haavelmo’s adoption of the F-N-P viewpoint that ‘stochasticity’ is intrinsic with data, but on several other methodological issues that pertain to the reliability of inference.
In current econometrics Haavelmo’s contributions seem largely irrelevant because there is no adequate appreciation or any concerted effort to address the methodological issues raised by seeking to bridge the gap between theory and data. What is esteemed is the mathematical sophistication of the inferential techniques (estimators and tests) called for in quantifying alternative theoretical models using different types of data (time series, cross-section and panel). Since the early 1950s, prestigious econometric journals are overflowing with a bewildering plethora of estimation techniques and their associated asymptotic theory focusing exclusively on addressing the ‘quantification’ problem, notwithstanding the gap between the variables envisaged by theory and what the available data measure. Worse, the emphasis is placed on mathematical derivations whose premises invoke mathematically convenient assumptions that are often non-testable and when they are testable they are rarely substantiated vis-a-vis the data. As a result, the incessant accumulation of asymptotic inferential results leave the practitioner none the wiser as to ‘how’ and ‘when’ to apply them to practical modeling problems with a view to learn from data about economic phenomena of interest. All the practitioners can do is apply these techniques to a variety of data hoping that occasionally some of the computer outputs will enable them to ‘tell a story’ and thus publish to survive academia. Hence, at the dawn of the 21st century, the applied econometric literature is filled with a disorderly assemblage of largely untrustworthy evidence that collectively shed no real light on economic phenomena of interest, primarily because the link between the theory and data is precarious at best. How did we reach this state of affairs?

A crucial implication of the textbook econometric pre-eminence of theory perspective is that by making probabilistic assumptions about error (shock) terms the emphasis is placed on the least restrictive assumptions that would justify (asymptotically) the ‘quantification’ technique, irrespective of whether they are testable or not. The idea is that weaker assumptions are less vulnerable to misspecification; hence the current popularity of the Generalized Method of Moments (GMM) as well as nonparametric methods. This is a flawed argument because weaker but indirect assumptions about the stochastic process \( \{Z_t, t \in \mathbb{N}\} \) underlying the data \( Z_0 \), that ignore the probabilistic structure of the data, are not necessarily less vulnerable to statistical misspecification, but they will invariably give rise to less precise inferences. In addition, non-testable assumptions made for mathematical convenience render the substantiation of statistical adequacy impossible. That is, the focus on ‘quantifying theoretical relationships’ subordinates the statistical premises to the justification of the estimation method employed, and invariably leads to unreliable inferences emanating from statistically misspecified models (statistically spurious results).

On this issue the textbook perspective is in direct conflict with one of Haavelmo’s key insights concerning the reliability and precision of inference based on a statistical model \( M_0(z) = \{f(z; \theta), \theta \in \Theta\}, \Theta \subset \mathbb{R}^m, z \in \mathbb{R}^n; \)

[B] the importance of parsimonious, but statistically adequate, (paramet-
ric) statistical models in ensuring the reliability and precision of inference:

"In all practical cases it is, therefore, necessary to be able to restrict, in advance, the set of admissible hypotheses Θ as much as possible, having at the same time strong reasons to believe that the true hypothesis is not outside this Θ."

In contrast to Haavelmo’s vision of building a bridge between theory and data, the primary focus of traditional textbook econometrics is on the ‘quantification of theory models assumed to be true’, largely ignoring the gap between theory and data Z0, as well as the link between the probabilistic assumptions made for the error terms and the structure of data Z0. Undoubtedly, this continues to contribute significantly to the untrustworthiness of empirical evidence produced by the textbook approach.

4.1 Revisiting the Simultaneous Equations model

To illustrate some of the issues raised above, consider the two equation structural (simultaneous equations) model:

\[
y_{1t} = \gamma_{12}y_{2t} + \delta_{11}x_{1t} + \delta_{12}x_{2t} + \delta_{13}x_{3t} + \varepsilon_{1t},
\]
\[
y_{2t} = \gamma_{21}y_{1t} + \delta_{21}x_{1t} + \delta_{23}x_{3t} + \delta_{25}x_{5t} + \varepsilon_{2t},
\]
whose generic Structural Form (SF) is:

\[
\Gamma^\top y_t + \Delta^\top x_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega),
\]

with the coefficient matrices Γ and Δ taking the form:

\[
\Gamma^\top := \begin{pmatrix} 1 & -\gamma_{12} \\ -\gamma_{21} & 1 \end{pmatrix}, \quad \Delta^\top := \begin{pmatrix} \delta_{11} & \delta_{12} & \delta_{13} & 0 & 0 \\ \delta_{21} & 0 & 0 & \delta_{24} & \delta_{25} \end{pmatrix}.
\]

For simplicity let us ignore the covariance Ω and consider the structural parameters \(\varphi := (\gamma_{12}, \gamma_{21}, \delta_{11}, \delta_{12}, \delta_{13}, \delta_{21}, \delta_{24}, \delta_{25})\), as the focus of the textbook quantification. Corresponding to this, there is a Reduced Form (RF):

\[
y_{1t} = \beta_{11}x_{1t} + \beta_{12}x_{2t} + \beta_{13}x_{3t} + \beta_{14}x_{4t} + \beta_{15}x_{5t} + u_{1t}
\]
\[
y_{2t} = \beta_{21}x_{1t} + \beta_{22}x_{2t} + \beta_{23}x_{3t} + \beta_{24}x_{4t} + \beta_{25}x_{5t} + u_{2t}
\]

derived by "solving" (2) for \((y_{1t}, y_{2t})\), which is specified in terms of the reduced form parameters: \(\theta := (\beta_{ij}, i = 1, 2, j = 1, 2, ..., 5)\). The general formulation of the RF, corresponding to the SF (3) is:

\[
y_t = B^\top x_t + u_t, \quad u_t \sim N(0, \Sigma).
\]

where \(\Gamma^\top y_t + \Delta^\top x_t = \varepsilon_t\) and \(y_t = B^\top x_t + u_t\) are related via \(B = (\Gamma^\top)^{-1} \Delta^\top\) and \(u_t = (\Gamma^\top)^{-1} \varepsilon_t\), giving rise to the identifying restrictions:

\[
B\Gamma + \Delta = 0, \quad \Omega = \Gamma^\top \Sigma \Gamma.
\]

Ignoring (for simplicity) any restrictions on Ω, the identifying restrictions among the structural and reduced form parameters take the implicit form:

\[
G(\varphi, \theta) = 0.
\]
where the structural parameters $\varphi := (\Gamma, \Delta)$ are said to be identified, if, for a given $\theta := B$, there exists a unique solution of (8) for $\varphi$. The identifying restrictions in the case of the structural model (2) are:

$$
\begin{align*}
\beta_{11} &= \beta_{21}\gamma_{12} + \delta_{11}, \\
\beta_{12} &= \beta_{23}\gamma_{12} + \delta_{12}, \\
\beta_{13} &= \beta_{22}\gamma_{12} + \delta_{13}, \\
\beta_{14} &= \beta_{24}\gamma_{12} + \delta_{14}, \\
\beta_{15} &= \beta_{25}\gamma_{12} \\
\beta_{21} &= \beta_{11}\gamma_{21} + \delta_{21}, \\
\beta_{24} &= \beta_{14}\gamma_{21} + \delta_{24}, \\
\beta_{25} &= \beta_{15}\gamma_{21} + \delta_{25}, \\
\beta_{22} &= \beta_{12}\gamma_{21}, \\
\beta_{23} &= \beta_{13}\gamma_{21}
\end{align*}
$$

(9)

In textbook econometrics, Haavelmo (1943) is, sometimes, credited with pointing out that when least-squares is applied to the structural model (2) it yields inconsistent estimators since:

$$
\text{Cov}(y_{2t}, \epsilon_{1t} | x_{1t}, x_{2t}, x_{3t}) \neq 0, \quad \text{Cov}(y_{1t}, \epsilon_{2t} | x_{1t}, x_{4t}, x_{5t}) \neq 0.
$$

To address the problem he propose the use of the method of Maximum Likelihood (ML) that gives rise to consistent and parameterization invariant estimators of $\varphi$. The latter property implies that ML is equivalent to using the MLE of $\theta$, $\hat{\theta}_{\text{MLE}}$, to derive the MLE of $\varphi$ via the restrictions (9):

$$
G(\hat{\varphi}_{\text{MLE}}; \hat{\theta}_{\text{MLE}}) = 0.
$$

That is, $\hat{\varphi}_{\text{MLE}}$ is the unique solution of (10), which, under certain probabilistic assumptions and regularity conditions, can be shown to be a Consistent and Asymptotically Normal (CAN) estimator of the structural parameters $\varphi$. What is often neglected in traditional econometrics is to validate, vis-a-vis data $Z_0$, the probabilistic assumptions invoked in deriving these optimal asymptotic properties of these estimators.

What is particularly interesting is that some of the methodological problems that contribute to the mountains of untrustworthy evidence could have been avoided if only textbook econometrics paid more attentions to Haavelmo (1944).

In addition to ignoring Haavelmo’s insistence that all inference should be based on the joint distribution $D(Z_1, Z_2, \ldots, Z_n; \theta)$ of the observable random variables $Z_t := (x_t, y_t)$, the traditional textbook econometrics also ignored the idea that the reduced form (5) is not just a subsidiary formulation with respect to which identification is considered, but it’s the statistical model in the context of which the structural model (2) is embedded. That is, textbook econometrics appears to ignore the fact that:

[C] behind every substantive (structural) model $M_{\varphi}(z)$, $z \in \mathbb{R}_+^m$, $\varphi \in \Phi \subset \mathbb{R}^m$ there is a statistical model $M_\theta(z)$, $z \in \mathbb{R}_+^n$, (often implicit) that pertains to the probabilistic structure of the observable stochastic process $\{Z_t, t \in \mathbb{N}\}$ underlying the data $Z_0$ in question (e.g. the textbook reduced form), and

[D] the reliability of the statistical ‘tools’ used to distinguish between ‘good’ and ‘bad’ models depend crucially on the validation — vis-a-vis data $Z_0$ — of the statistical model $M_\theta(z)$ implicit in $M_{\varphi}(z)$.

When any of the probabilistic assumptions comprising $M_\theta(z)$ are invalid for data $Z_0$, the inferential tools, which include not only point estimates, standard errors and significance testing results, but also goodness-of-fit/prediction measures, are rendered
unreliable. Indeed, statistical misspecification calls serious into question the ‘quantification of theory’ by rendering any inference based on the estimated model unreliable; see Spanos (1990). As Haavelmo (1944) rhetorically asked:

“What is the use of testing, say, the significance of regression coefficients, when maybe, the whole assumption of the linear regression equation is wrong?” (p. 66)

On this issue of a well-defined statistical model, he was highly insistent:

“... economists might get more useful and reliable information (and also fewer spurious results) out of their data by adopting more clearly formulated probabilistic models; and that such formulation might help in suggesting what data to look for and how to collect them.” (p. 114)

This implicit statistical model \( M_\phi(z) \) behind the structural model \( M_{\varphi}(z) \) in (3), represented by (6), takes the more explicit form of the Multivariate Normal, Linear Regression model (table 1), resulting from a particular parameterization of the vector stochastic process \( \{Z_t, t \in \mathbb{N}\} \) underlying the data \( Z_0 \) in question; see Spanos (2006b). The probabilistic assumptions [1]-[5] are specified in terms of the observables and represent a complete and internally consistent set of assumptions whose validity vis-a-vis the data \( Z_0 \) establishes the statistical adequacy of \( M_\theta(z) \). The key difference between \( M_\theta(z) \) in table 1 and the reduced form (6) is that everything in table 1, including the functional forms, stem from and are justified by the probabilistic structure of \( \{Z_t, t \in \mathbb{N}\} \). This secures the reliability of any inductive inferences based on the latter by ensuring that the actual error probabilities are close enough to the nominal (assumed) ones.

### Table 1 - Multivariate Normal, Linear Regression Model

<table>
<thead>
<tr>
<th>Statistical GM</th>
<th>( y_t = \beta_0 + B_1^T x_t + u_t, \ t \in \mathbb{N} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Normality</td>
<td>( (y_t</td>
</tr>
<tr>
<td>[2] Linearity</td>
<td>( E(y_t</td>
</tr>
<tr>
<td>[3] Homosk/city</td>
<td>( Var(y_t</td>
</tr>
<tr>
<td>[4] Independence</td>
<td>( { (y_t</td>
</tr>
<tr>
<td>[5] t-invariance</td>
<td>( \theta := (\beta_0, B_1, \Sigma) ) do not change with ( t ).</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\beta_0 &= E(y_t) - B_1^T E(X_t) \\
B_1 &= \text{Cov}(X_t)^{-1} \text{Cov}(X_t, y_t) \\
\Sigma &= \text{Cov}(y_t) - \text{Cov}(y_t, X_t) B_1
\end{align*}
\]

The current practice in textbook econometrics often ignores the RF and its adequacy and estimates the structural model directly, seemingly oblivious to the fact that the statistical properties attributed to the estimators and tests relating to the structural parameters are likely to be invalid when the RF is statistically misspecified; Spanos (1986, 1990). Indeed, the overwhelming majority of published papers using the simultaneous equations model sidestep the statistical model in table 1 and go directly to the ‘quantification’ of the structural model; see Spanos (2006a).

The crucial step that is routinely ignored by textbook econometrics is to validate the statistical model \( M_\theta(z) \) (test assumptions [1]-[5] in table 1) and establish a
sound link between data \( \mathbf{Z}_0 \) and the structural model \( \mathcal{M}_\varphi(\mathbf{z}) \) in (3), using thorough Mis-Specification (M-S) testing (see Spanos, 1986), before any inferences are drawn. This step is also implicit when one is using Instrumental Variables and other related estimation methods like the Generalized Method of Moments; see Spanos (2007b). One will be very hard pressed to find a handful of applied papers in over 100,000 of published papers in all prestigious journals in economics over the last 70 years that did a good job with this statistical adequacy step.

The statistical adequacy of \( \mathcal{M}_\theta(\mathbf{z}) \) constitutes a precondition for ensuring the reliability of any inference pertaining to the structural parameters \( \varphi \), including the validity of the ‘testable’ substantive information in \( \mathcal{M}_\varphi(\mathbf{z}) \) based on testing the overidentifying restrictions that can be framed in terms of the hypotheses:

\[
H_0 : \mathbf{G}(\varphi, \theta) = \mathbf{0}, \text{ vs. } H_1 : \mathbf{G}(\varphi, \theta) \neq \mathbf{0}. \tag{11}
\]

When \( \mathcal{M}_\theta(\mathbf{z}) \) is statistically misspecified, not only is the identification of \( \varphi \) questionable, but the results of such a test will be misleading because the relevant actual error probabilities will invariably be significantly different from the nominal ones. In the above example, there are 2 overidentifying restrictions in (4) because there are 10 statistical parameters in (5) but only 8 structural parameters in (2); see Spanos (2005).

A structural model \( \Gamma(\varphi)^\top \mathbf{y}_t + \Delta(\varphi)^\top \mathbf{x}_t = \varepsilon_t \), is said to be data-pertinent when:

(a) the associated statistical model \( \mathbf{y}_t = \mathbf{B}^\top(\theta)\mathbf{x}_t + \mathbf{u}_t \), is statistically adequate and

(b) the overidentifying restrictions: \( \mathbf{G}(\varphi, \theta) = \mathbf{0} \) are data-acceptable.

Under (a)-(b) the estimated empirical model:

\[
\mathcal{M}_\varphi(\mathbf{z}) : \Gamma(\hat{\varphi})^\top \mathbf{y}_t + \Delta(\hat{\varphi})^\top \mathbf{x}_t = \hat{\varepsilon}_t, \tag{12}
\]

enjoys both (i) statistical adequacy and (ii) theoretical meaningfulness, and can be used for prediction and policy simulations as well as a basis for learning from data about the phenomenon of interest.

As argued next, the number of Haavelmo’s insights, whose disregard contributed significantly to the untrustworthiness of current empirical evidence in economics, is not limited to [A]-[D] above.

5 Beyond the Fisher-Neyman-Pearson perspective

The methodological framework envisaged by Haavelmo (1944) went beyond the F-N-P perspective in an obvious attempt to account for the fact that economic data do not arise from designed experiments but from passive observation:

“... such [economic] research has to build, mostly, on passive observations of facts, instead of data obtained by rationally planned experiments (see Chapter II). And this means that we can obtain only such data as are the results of the economic system as it in fact is, and not as it would be under those unrestricted hypothetical variations with which we operate in economics theory.” (p. 85)

In light of that, Haavelmo viewed the problem of modeling with observational data as one of bridging the gap between theory and data.
“The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier. But the bridge itself was never completely built.” (p. iii)

The bridge pier he envisaged was based on drawing a number of important distinctions, including:

[E] distinguishing between theoretical, observational, true variables and the observed data, emphasizing that:

“When we set up a system of theoretical relationships and use economic names for the otherwise purely theoretical variables involved, we have in mind some actual experiment, or some design of experiment, which we could at least imagine arranging, in order to measure those quantities in real economic life that we think might obey the laws imposed on their theoretical namesakes.” (p. 6)

“... he [the economist] is presented with some results which, so to speak, Nature has produced in all their complexity, his task being to build models that explain what has been observed.” (p. 7)

“... one should study very carefully the actual series considered and the conditions under which they were produced, before identifying them with the variables of a particular theoretical model.” (p. 7)

Haavelmo understood that one needs, somehow, to bridge the gap between theory and data by ‘adjusting’ one of the two:

"In order to test a theory against facts, or to use it for predictions, either the statistical observations available have to be "corrected", or the theory itself has to be adjusted, so as to make the facts we consider the "true" variables relevant to the theory.” (p. 7)

With that in mind, Haavelmo (1944) suggested that in confronting economic theories with data one needs to specify the circumstances envisaged by the theory (what he called a design of experiments) that link the two:

[F] distinguishing between designs based on artificial isolation and the designs of Nature (p. 14):

“(1) experiments that we should like to make to see if certain real phenomena — when artificially isolated from "other influences" — would verify certain hypotheses. and (2) the stream of experiments that Nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers. In both cases the aim of theory is the same, namely, to become master of the happenings of real life. But our approach is a little different in the two cases. ...

In the second case we can only try to adjust our theories to reality as it appears before us. ... We try to choose a theory and a design of experiments to go with it, in such a way that the resulting data would be those which we get by passive observation of reality.”

Haavelmo went on to unveil a particularly crucial problem in bridging the gap between theory and data, arguing that:

[G] economic theories often invoke artificial isolation experiments, but economic data result from designs of Nature, and warned:

“... theories are often being compared with data which cannot at all be considered
as observations obtained by following the design of experiments we had in mind when constructing the theory.” (p. 15)

With his first type of design of experiments Haavelmo (1944) anticipated modern experimental economics (see Smith, 2007) and posed certain key questions that needed to be asked by the modeler:

“(c) Why do we confine ourselves only to such theories as are directly verifiable? Or, why are we interested in relations for which Nature does not furnish experiments?

(d) Very often our theories are such that we think certain directly observable series would give adequate experimental results for verification, provided other things did not change. ...

(e) Are we interested in describing what actually does happen, or are we interested in what would happen if we could keep "other things" unchanged? " (p. 16)

With his second type of experiments Haavelmo anticipated the endemic untrustworthiness of current empirical evidence in applied econometrics when theories are confronted with the ‘wrong’ data. Haavelmo (1944) warned econometricians about the dangers waiting the unaware if a number of key questions are not posed and addressed adequately:

“(a) Are most of the theories we construct in "rational economics" one for which historical data and passive observations are not adequate experiments? This question is connected with the following:

(b) Do we try to construct theories describing what individuals, firms, etc. actually do in the course of events, or do we construct theories describing schedules of alternatives at a given moment? If the latter is the case, what bearing do such schedules of alternatives have upon a series of decisions and actions actually carried out? (p. 16)

One can only imagine how empirical evidence in economics would have been a lot more trustworthy, were Haavelmo’s warning taken seriously by current econometric practice. One only needs to glance at the current practice in any area of applied econometrics to realize that none of his insights had any real impact.

In concluding this section it is important to emphasize that Haavelmo (1944) includes numerous additional methodological ideas and insights in addition to [A]-[G] discussed above. Just to mention a few:

[H] the distinction between potential and factual influencing factors,
[I] the distinction between the constancy of the estimated relationships and the autonomy of the underlying theoretical relationships,
[J] the importance of a causal structure in evaluating the adequacy of substantive models for explaining the observable phenomena of interest,
[K] the philosophical problem of underdetermination of theory from data (p. 74), and
[L] the legitimacy of exploratory data analysis in assessing the validity of a statistical model $M_0(z)$.

Some of these ideas and insights will play an important role in the next section where an attempt will be made to build the bridge pier Haavelmo (1944) envisaged.
6 On completing Haavelmo’s *bridge pier*

The main argument so far can be summarized as an attempt to bring out some of Haavelmo’s key methodological ideas and insights in order to make a case that the failure to integrate them into the current econometric practice, as articulated by textbook econometrics, contributed significantly to the untrustworthiness of the published empirical evidence going back more than half a century.

During the years that Haavelmo was working on the 1944 monograph (see Bjerkholt, 2007), the probabilistic perspective he had in mind is best described in Neyman’s 1937/1952 book (cited in Haavelmo, 1944, p. 77):

“The application of the theory [of probability] involves the following steps:

(i) If we wish to treat certain phenomena by means of the theory of probability we must find some element of these phenomena that could be considered as random, following the law of large numbers. This involves a construction of a mathematical model of the phenomena involving one of more probability sets.

(ii) The mathematical model is found satisfactory, or not. This must be checked by observation.

(iii) If the mathematical model is found satisfactory, then it may be used for deductions concerning phenomena to be observed in the future.” (Neyman, 1952, p. 27)

In this quotation Neyman demarcates in (i) the domain of statistical modeling to include stochastic (random) phenomena: observed phenomena that exhibit enough chance regularity patterns to render the Law of Large Numbers (LLN) applicable. In this sense the LLN demarcates the limits of statistical modeling in the sense that it specifies the weakest probabilistic assumptions on the process \( \{Z_t, t \in \mathbb{N}\} \) that would give rise to a potentially operational statistical model. This brings out the reliance of frequentist inductive inference on the long-run constancy of relative frequencies. Like Fisher, Neyman in (ii) emphasizes the testing of the assumptions comprising the statistical model in order to ensure its adequacy. In (iii) he clearly indicates that statistical adequacy is a necessary condition for any inductive inference. This is because the ‘error probabilities’, in terms of which the optimality of inference is defined, depend crucially on the validity of the model:

“... any statement regarding the performance of a statistical test depends upon the postulate that the observable random variables are random variables and posses the properties specified in the definition of the set \( \Omega \) of the admissible simple hypotheses.” (Neyman, 1950, p. 289)

A crucial implication of this is that when the statistical model is misspecified, the actual error probabilities, in terms of which ‘optimal’ inference procedures are chosen, are likely to be very different from the nominal ones, leading to unreliable inferences.

The question that naturally arises is whether one can integrate Haavelmo’s methodological ideas and insights into a broader modeling framework in order to address the endemic untrustworthiness of evidence problem. The initial attempt to do just that was made in Spanos (1986; 1988; 1989) where some of Haavelmo’s ideas and insights were integrated into the Fisher-Neyman-Pearson probabilistic perspective.
This initial attempt is schematically shown in the diagram below.

![Diagram 1: A proposed framework for empirical modeling](image)

A key feature of the proposed framework, the distinction between a theory model $\mathcal{M}_\psi(z, \xi)$, a structural model $\mathcal{M}_\varphi(z)$ and the statistical model $\mathcal{M}_\theta(z)$, was inspired by Haavelmo’s distinction between theoretical, observational, true variables and the observed data mentioned in [F] above.

Relating theory to data goes beyond the Fisher-Neyman-Pearson inferential statistics, and calls for an explicit modeling stage that links the two using a bridge based on several interconnected models.

From the theory side of the proposed bridge one constructs a theory model (a mathematical formulation of the theory in question), say $\mathcal{M}_\psi(z; \xi)$, which often includes latent variables $\xi$. Any attempt to confront $\mathcal{M}_\psi(z; \xi)$ with data $Z_0$, however, calls for a metamorphosis of $\mathcal{M}_\psi(z; \xi)$ into an estimable (in light of data $Z_0$) form, called the structural (substantive) model $\mathcal{M}_\varphi(z)$. The construction of a structural model is particularly difficult in economics because one needs to reflect on the huge gap between the circumstances envisaged by the theory and its concepts (intentions, plans), and the actual phenomenon of interest giving rise to the available data; see Spanos (1995).

From the data side, the statistical information (chance regularity patterns exhibited by data) is ‘accounted for’ by selecting a statistical model $\mathcal{M}_\theta(z)$ with a view to meet two interrelated aims:

(i) to adequately account for the chance regularities in data $Z_0$ (Spanos, 1999), by choosing a probabilistic structure for the stochastic process $\{Z_t, t \in \mathbb{N}\}$ so as to render $Z_0$ a ‘truly typical realization’ thereof, and
(ii) to parameterize the chosen probabilistic structure in an attempt to specify \( \mathcal{M}_\theta(z) \) in such a way so as to embed (parametrically) \( \mathcal{M}_\varphi(z) \) in its context via general restrictions of the form \( G(\theta, \varphi) = 0 \), \( \theta \in \Theta \), \( \varphi \in \Phi \), relating the statistical \( (\theta \in \Theta) \) and substantive \( (\varphi \in \Phi) \) parameters. This view of a statistical model is related to Haavelmo’s view of statistical models as parameterizations stemming from the joint distribution \( D(Z_1, Z_2, \ldots, Z_n; \phi) \).

To assess the appropriateness of the selected statistical model one needs to test its adequacy vis-a-vis data \( Z_0 \) using thorough M-S testing, and if found wanting, to be respecified until a statistically adequate model \( \mathcal{M}_\theta(z) \) is found that parametrically nests \( \mathcal{M}_\varphi(z) \); see Spanos (2010). Only then should one consider the identification problem relating to solving the implicit restrictions \( G(\theta, \varphi) = 0 \) uniquely for \( \varphi \), as well as employing statistical procedures to seek reliable answers to substantive questions of interest, including testing the overidentifying restrictions in (11). In an important sense a statistically adequate model \( \mathcal{M}_\theta(z) \) transforms a finite and incomplete set of raw data \( z_0 \) – containing uncertainties, impurities and noise – into reliable ‘evidence’ to confront \( \mathcal{M}_\varphi(z) \) and other substantive hypotheses of interest with.

Like the bridge pier envisaged by Haavelmo (1944), the modeling framework proposed in Spanos (1986) was found to be incomplete primarily because the Fisher-Neyman-Pearson framework left several key foundational problems unresolved, including model validation:

“\textit{The current statistical methodology is mostly model-based, without any specific rules for model selection or validating a specified model.}" (C. R. Rao, 2004, p. 2)

To get some idea of the extent of the task to complete the envisaged modeling framework the key problems that needed to be addressed are listed below in two sub-categories.

\textbf{A. Foundational issues pertaining to Modeling}

[i] \textbf{Statistical model specification:} how to select the prespecified \( \mathcal{M}_\theta(z) \),

[ii] \textbf{the role of statistical information} contained in data \( Z_0 \) in the specification of \( \mathcal{M}_\theta(z) \) (Spanos, 2006b),

[iii] \textbf{the role of substantive (subject matter) information} in statistical modeling (Lehmann, 1990, Cox, 1990),

[iv] \textbf{the nature, structure and role of the notion of a statistical model} \( \mathcal{M}_\theta(z), \ z \in \mathbb{R}_n^\mathbb{Z} \),

[v] \textbf{the role of the data in guiding model discovery,} i.e. the legitimacy of \textit{exploratory data analysis} (e.g. data mining, double-use of data, pre-test bias, circularity and infinite regress) (Spanos, 2000),

[vi] \textbf{statistical adequacy:} how to establish the validity of a statistical model \( \mathcal{M}_\theta(z) \) \textit{a posteriori},

[vii] \textbf{statistical model re-specification:} how to respecify a model \( \mathcal{M}_\theta(z) \) when found misspecified, and

[viii] \textbf{substantive adequacy:} how to assess the adequacy of a substantive model \( \mathcal{M}_\varphi(z) \) in elucidating (describe, explain, predict) the phenomenon of interest.
B. Foundational issues pertaining to Inference

[ix] the role of pre-data vs. post-data error probabilities (Savage, 1962; Hacking, 1965),
[x] safeguarding frequentist inference against:
(a) the fallacy of acceptance: (mis)interpreting accept \( H_0 \) [no evidence against \( H_0 \)] as evidence for \( H_0 \); e.g. the test failed to detect an existing discrepancy because it had low power,
(b) the fallacy of rejection: (mis)interpreting reject \( H_0 \) [evidence against \( H_0 \)] as evidence for a particular \( H_1 \); e.g. conflating statistical with substantive significance (Mayo, 1996; Mayo and Spanos, 2010, 2011),
[xi] a sound frequentist interpretation of probability that provides an adequate foundation for frequentist modeling and inference (Spanos, 2011b),
[xii] parametric vs. nonparametric modeling and inference, and

The present paper focuses almost exclusively on issues [i]-[viii], paying particular attention to model validation that secures the error reliability of inductive inference. For extensive discussions pertaining to inference problems, see Mayo and Cox (2006), Mayo and Spanos (2004, 2006) and Spanos (2000, 2001, 2007a, 2010a, 2011b). These papers are relevant for the discussion that follows because, when taken together, they demarcate what Mayo (1996) called the ‘error statistical approach’ that offers a unifying inductive reasoning for frequentist inference. In this sense, error statistics can be viewed as a refinement/extension of the Fisher-Neyman-Pearson (F-N-P) motivated by the call to address the above foundational problems; Mayo and Spanos (2011). In particular, error statistics aims to:

[A] refine the F-N-P approach by proposing a broader framework with a view to secure statistical adequacy, motivated by the foundational problems [i]-[vii] (Spanos, 2011a), and

[B] extend the F-N-P approach by supplementing it with a post-data severity assessment with a view to address problems [ix]-[xiii] (Mayo and Spanos, 2006).

The refinement/extension of the F-N-P approach was necessary to address some of the key foundational problems bedeviling frequentist inference since the 1940s. In error statistics probability plays two interrelated roles. Firstly, \( f(z; \theta) \), \( z \in \mathbb{R}_z^m \) attributes probabilities to all legitimate events relating to the sample \( Z \). Secondly, it provides all relevant error probabilities associated with any statistic (estimator, test or predictor) \( T_n = g(Z) \) via: \( F(t; \theta) = \mathbb{P}(T_n \leq t; \theta) = \int \cdots \int f(z; \theta)dz \).

Pre-data these error probabilities quantify the generic capacity of any inference procedures to discriminate among alternative hypotheses, framed in terms of an unknown parameter \( \theta \). Post-data error probabilities are used to establish the warranted discrepancies \( \gamma \) from prespecified values of \( \theta \), using the severity assessment. For the error statistician probability arises, post-data, not to measure degrees of confirmation.
or belief in hypotheses, but to quantify the extent to which a given data \( Z_0 \) provide evidence for or against a particular hypothesis. There is evidence for a particular statistical hypothesis or claim just to the extent that the test that passes such a claim with \( Z_0 \) is severe: that with high probability the hypothesis would not have passed so well as it did if it were false, or specific departures were present; see Mayo and Spanos (2011).

A key methodological problem that has bedeviled every applied field that relies on observational data relates to the role of theory and data in empirical modeling and pertains to the issues [i]-[iv] raised above. The pre-eminence of theory perspective that has dominated empirical modeling in economics since Ricardo (1817) is that the inevitable result of foisting one’s favorite theory on the data is often an estimated model which is both statistically and substantively misspecified, but one has no principled way to distinguish between the two sources of misspecification and apportion blame:

\[
is\text{ the substantive subject matter information false? or are the statistical premises of inference misspecified? (13)}
\]

The key to circumventing this Duhemian ambiguity is to find a way to disentangle the statistical \( \mathcal{M}_\theta(z) \) from the substantive premises \( \mathcal{M}_\phi(z) \), without compromising the integrity of either source of information. It is well-known that \( \mathcal{M}_\phi(z) \) comes from the theory model in light of data \( Z_0 \). But where does the statistical model \( \mathcal{M}_\theta(z) \) come from? In the context of the simultaneous equations model, the textbook econometrics perspective presumes that it should come from \( \mathcal{M}_\phi(z) \), being its reduced form. But that’s where the key problem lies: the probabilistic structure is implicitly determined by the structural model \( \mathcal{M}_\phi(z) \), without any attempt to account for the chance regularities in data \( Z_0 \), giving rise to the Duhemian ambiguity in (13).

6.1 Where do statistical models come from?

The gist of how to disentangle the statistical from the substantive information lies in Haavelmo’s view of statistical models as stemming from the joint distribution of the observables involved.

The construction of the statistical model (premises) \( \mathcal{M}_\theta(z) \) begins with a given data \( Z_0 \), irrespective of the theory or theories that led to the choice of \( Z_0 \). Once selected, data \( Z_0 \) take on ‘a life of their own’ as a particular realization of an underlying generic stochastic process \( \{ Z_t, t \in \mathbb{N} \} \). The link between data \( Z_0 \) and the process \( \{ Z_t, t \in \mathbb{N} \} \) is provided by the key question:

\[
\text{‘what probabilistic structure pertaining to the process } \{ Z_t, t \in \mathbb{N} \} \text{ would render data } Z_0 \text{ a truly typical realization thereof?’}
\]

A pertinent answer to this question provides the relevant probabilistic structure for \( \{ Z_t, t \in \mathbb{N} \} \). Given that, one then chooses a parameterization \( \theta \in \Theta \) for \( \{ Z_t, t \in \mathbb{N} \} \) to define the relevant statistical model \( \mathcal{M}_\theta(z) \) that parametrically nests the substantive model \( \mathcal{M}_\phi(z) \).
What does a ‘typical realization’ of Normal, Independent and Identically Distributed (NIID) process \( \{Z_t, \, t \in \mathbb{N}\} \) look like? Like fig. 1, not 2-4!

**Example 1.** For the data in fig. 1, an appropriate model is the simple Normal:

\[ M_{\theta}(z): Z_t \sim \text{NIID}(\mu, \sigma^2), \quad \theta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+, \quad z_t \in \mathbb{R}, \quad t \in \mathbb{N} \]

**Example 2.** For fig. 2, an appropriate model is the simple Exponential:

\[ M_{\theta}(z): Z_t \sim \text{Weibull-IID}(\mu, \delta), \quad \theta := (\mu, \delta) \in \mathbb{R}^2_+, \quad z_t \in \mathbb{R}_+, \quad t \in \mathbb{N} \]

**Example 3.** For fig. 3, an appropriate model is the Normal, mean-heterogeneous:

\[ M_{\theta}(z): Z_t \sim \text{N}(\mu_0 + \mu_1 t, \sigma^2), \quad \theta := (\mu_0, \mu_1, \sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+, \quad z_t \in \mathbb{R}, \quad t \in \mathbb{N} \]

**Example 4.** For fig. 4, an appropriate model is the Normal AR(1):

\[ M_{\theta}(z): (Z_t | Z_{t-1}) \sim \text{N}(\alpha_0 + \alpha_1 Z_{t-1}, \sigma_0^2), \quad \theta := (\alpha_0, \alpha_1, \sigma_0^2) \in \mathbb{R} \times (-1, 1) \times \mathbb{R}_+, \quad t \in \mathbb{N} \]

This purely statistical perspective that enables one to separate the ‘statistical’ from the ‘substantive’ information is analogous to Shannon’s (1948) information theory which is based on formalizing the informational content of a message by separating ‘regularity patterns’ in strings of ‘bits’ from any substantive ‘meaning’: “Frequently
the messages have meaning ... These semantic aspects of communication are irrelevant to the engineering problem."

Analogously, the statistical perspective formalizes statistical information in terms of probability theory, e.g. probabilistic assumptions pertaining to the process \( \{Z_t, t \in \mathbb{N}\} \) underlying data \( Z_0 \), and the substantive information (meaning) is irrelevant to the purely statistical problem of validating the statistical premises.

The distinction is particularly important in delineating between:

[a] **statistical adequacy**: does \( M_\theta(z) \) account for the chance regularities in \( Z_0 \)?

[b] **substantive adequacy**: does the model \( M_\varphi(z) \) shed adequate light on (describe, explain, predict) the phenomenon of interest?

The two are related in so far as the former is necessary, but not sufficient, for securing the latter. To establish substantive adequacy one needs to probe for additional potential errors vis-a-vis the phenomenon of interest, like overly narrow demarcation of the phenomenon of interest, impractical *ceteris paribus* clauses, intentions vs. realizations, omitted but relevant effects, false causal claims, etc. This means that the potential errors one needs to probe for and guard against are very different in the two cases. One of the key advantages of bridging the gap between theory and data in terms of a sequence of interrelated models (diagram 1) is that it makes the framing of potential errors during the modeling process more transparent; see Spanos (2012).

### 6.2 The importance of Statistical vs. substantive adequacy

This distinction can be used to shed light on several confusions that permeate the current textbook approach to econometric modeling, including the following.

1. One often encounters empirical modelers who justify their neglect of statistical adequacy by repeating the slogan: "all models are false, but some are useful", attributed to George Box (of the Box-Jenkins fame). This claim is based on confusing substantive and statistical inadequacy. A **substantive model** \( M_\varphi(z) \) may always come up short in fully capturing or explaining a phenomenon of interest [e.g. not realistic enough], but a statistical model \( M_\theta(z) \) may be entirely satisfactory [when statistically adequate] to reliably test substantive questions of interest, including assessing the appropriateness of \( M_\varphi(z) \).

2. **Learning from data** depends crucially on establishing a **sound link** between the process generating data \( z_0 \) and the assumed \( M_\theta(z) \), by securing statistical adequacy that ensures the closeness of actual and nominal error probabilities. Without it, the inference results are just artifacts, that, when relied upon, can only shed misleading light on the phenomenon of interest.

3. **Statistical adequacy** depends only on \( M_\theta(z) \) and \( Z_0 \) and can be established independently by different practitioners using thorough M-S testing; it is well-defined and objective because everything is on the table. Hence, when performed by informed practitioners their M-S diagnosis and respecification should give rise to same respecified model. Hence, to paraphrase Keynes (1940):

> "if seventy well-trained econometricians were shut up in seventy separate rooms with..."
the same $M_{\phi}(z)$, $M_{\theta}(z)$ and $Z_0$, when they emerge they should have the same misspecification diagnosis and the same (adequate) respecified statistical model."

Such a claim cannot be made in relation to establishing substantive adequacy because the search is more open-ended in the sense that there is an infinity of scenarios one could potentially probe.

4. On one hand, it is insufficiently understood that **good fit/prediction** is neither necessary nor sufficient for the statistical adequacy of $M_{\theta}(z)$, but it can be relevant for the **substantive adequacy** of $M_{\phi}(z)$, in the sense that it provides a measure of its comprehensiveness (explanatory capacity) vis-a-vis the phenomenon of interest. One the other hand, it is often insufficiently appreciated that statistical inadequacy also undermines the reliability of such goodness-of-fit measures, and calls into question any claims pertaining to the ‘theoretical meaningfulness’ of the sign or magnitude of estimated parameters, however informal.

5. Statistical adequacy can be used to impugn the kind of "**spurious results**" like those unveiled by Yule (1926) and alluded to by Haavelmo (1944), p. 114, as stemming from **statistical misspecification**. Time series data that exhibit both temporal dependence and heterogeneity will give rise to a statistically misspecified linear regression model because invariably assumptions [2]-[5] (table 1) are likely to be invalid; see Spanos and McGuirk (2001). Worse, the mistaken belief that goodness-of-fit measures can be used to evaluate the extent of which the estimated model accounts for the regularities in the data has led to another misleading impression that one can ‘forge’ significant correlations (or regression coefficients) at will, if one was prepared to persevere long enough ‘mining’ the data. That led to yet another (mistaken) impression, that statistical spuriousness is both inevitable and endemic, especially when analyzing observational data; this impression is almost universal among social scientists and philosophers. This has led to widely held (but unwarranted) belief that substantive information provides the only safeguard against statistical spuriousness. As a result, the prevailing pre-eminence of theory perspective in economics had led to persistently ostracizing any attempt to redress the balance between theory and data, by charging any practices that strive for giving data an separate voice, with ‘measurement without theory’, ‘data-mining’ and ‘hunting’ for statistical significance.

6. In the social sciences, where **the actual DGM is highly complex** and the available **data are (often) observational**, a statistically adequate $M_{\theta}(z)$ could play a more crucial role in guiding the search for substantively adequate explanations (theories) by delineating ‘what there is to explain’. Kepler’s ‘law’ pertaining to the elliptical motion of the planets was an empirical regularity that guided Newton toward a substantively adequate model, that of universal gravitation, 60 years after the empirical regularity was first established; see Spanos (2007).

7. What about the notion of a **model is adequate for a purpose**? That notion pertains only to the substantive adequacy. Substantively simple (unrealistic) models can lead to learning from data about a phenomenon of interest only when they are statistically adequate; without statistical adequacy no reliable learning/inference is
possible, whatever the purpose.

8. The confusion between structural and statistical error terms. Returning to the simultaneous equations model in (3), the structural error term $\varepsilon_t$ is autonomous and global in the sense that it can represent errors of approximation, measurement errors, left out effects, etc. The statistical error term in table 1, however, is derived and local in the sense that, by definition, $u_t = y_t - E(y_t | x_t)$, and can only represent systematic statistical information already in data $Z_0$, but not accounted for by the assumed systematic component: $E(y_t | x_t) = \beta_0 + B_1^T x_t$.

9. The distinction brings out how confused is the textbook discussion of the ‘omitted variables’ bias and inconsistency problem. A closer look at the argument reveals that it has nothing to do with statistical misspecification; it is an issue pertaining to the substantive adequacy; Spanos (2006c).

10. The distinction between substantive vs. statistical adequacy elucidates the crucial differences between statistical model specification and model selection based on Akaike-type criteria; Spanos (2010b).

7 The textbook approach and the unreliability of inference: the case of DSGE modeling

Some of the key features of the early empirical modeling practices that Haavelmo (1944) called into question are still reigning strong over both the traditional econometric textbook discussions and applied econometrics research, including the Dynamic Stochastic General Equilibrium (DSGE) modeling that dominates current macroeconometrics; see Spanos (2009), Poudyal and Spanos (2011).

Dynamic Stochastic General Equilibrium (DSGE) models aim to describe the behavior of the economy in an equilibrium steady state by analyzing the interaction of microeconomic decisions between several agents (households, firms, governments, central banks); deterministic theory-models in the form of a system of first order difference equations driven by latent stochastic shocks. The basic DSGE model chosen is discussed in Ireland (2004) is given in table 2.

<table>
<thead>
<tr>
<th>Table 2: Theory Model: behavioral equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln(\frac{Y_t}{Q_t}) = \ln(\frac{Y_t}{Z_t}) - \omega \ln(a_t)$</td>
</tr>
<tr>
<td>$\ln(\frac{Y_t}{Y_{t-1}}) = \ln(\frac{Y_t}{Z_t}) - \ln(\frac{Y_{t-1}}{Z_{t-1}}) + \ln(\frac{Z_t}{Z_{t-1}})$</td>
</tr>
<tr>
<td>$\ln(\frac{Y_t}{Q_t})=\alpha_x \ln(\frac{Y_{t-1}/Q_{t-1}}{\bar{Y}<em>x}) + (1-\alpha_x)E_t \ln(\frac{Y</em>{t+1}/Q_{t+1}}{\bar{Y}_x}) - \left{ \ln \left( \frac{R_t}{\bar{R}<em>t} \right) - E_t \ln(\frac{P</em>{t+1}/P_t}{\bar{P}}) \right} + (1-\omega)(1-\rho_a) \ln(a_t)$</td>
</tr>
<tr>
<td>$\ln(\frac{P_t}{P_{t-1}})=\beta \left( \alpha_x \ln(\frac{P_{t-1}/P_{t-2}}{\bar{P}<em>x}) \right) + (1-\alpha_x)E_t \ln(\frac{P</em>{t+1}/P_t}{\bar{P}_x}) + \psi \ln(\frac{Y_t/Q_t}{\bar{Y}}) - (\frac{1}{\bar{P}_x}) \ln(\frac{\bar{R}_t}{\bar{P}_x})$</td>
</tr>
<tr>
<td>$\ln(\frac{R_t}{\bar{R}<em>t}) - \ln(\frac{R</em>{t-1}}{\bar{R}<em>{t-1}}) = \rho_x \ln(\frac{P</em>{t+1}/P_t}{\bar{P}<em>x}) + \rho_g \ln(\frac{Y_t/Q_t}{\bar{Y}}) + \rho_y \ln(\frac{Y_t/Q_t}{\bar{Y}}) + \varepsilon</em>{rt}$</td>
</tr>
</tbody>
</table>

DSGE theory-model $M_{\Phi}(z; \xi)$, is specified in terms of the following observable
variables $Z_t = (Y_t, C_t, P_t, R_t)$, where $Y_t$-production, $C_t$-consumption, $P_t$-price level, and $R_t$ - interest rate, as well as the latent variables $\xi = (Q_t, Z_t, a_t, \theta_t)$, where $Q_t$-efficient level of output, $X_t = (Y_t/Q_t)$-output gap, $p_t = (P_t/P_{t-1})$-temporal price ratio, shocks: $a_t$-preference, $\theta_t$-elasticity of demand, $Z_t$-technology.

Note that $\tau = \gamma = \frac{1}{n} \sum_{t=1}^{n} (Y_t/Y_{t-1}), \quad p = \frac{1}{n} \sum_{t=1}^{n} (P_t/P_{t-1}), \quad \tau = \frac{1}{n} \sum_{t=1}^{n} (X_t/X_{t-1})$.

Stochastic shock equations

$$\ln(a_t) = \rho_\alpha \ln(a_{t-1}) + \varepsilon_{at}, \quad (1/\phi) \tau_t = \rho_\theta (1/\phi) \theta_{t-1} + \varepsilon_{\theta t}, \quad \ln(Z_t/Z_{t-1}) = \varepsilon_{zt},$$

Structural parameters: $\omega, \alpha_x, \rho_\alpha, \beta, \psi, \alpha_x, \phi, \rho_\pi, \rho_\gamma, \rho_x, \rho_\theta, \omega = \frac{1}{\eta}, \psi = \frac{\eta(\theta-1)}{\phi}$

Calibration: $\tau = \gamma = 1.0048, \quad p = 1.0086, \quad \bar{R} = \frac{1}{T} \sum_{t=1}^{n} R_t = \frac{1}{T} \bar{F} \rightarrow \beta = .99, \quad \psi = .1$.

The transformations applied to the observables $Z_t = (Y_t, C_t, P_t, R_t)$ are clearly theory-oriented [attempt to impose the steady-state] but show no respect for the probabilistic structure of the data. For instance, why would one scale $Y_t/Y_{t-1}$ using $\gamma = \frac{1}{n} \sum_{t=1}^{n} (Y_t/Y_{t-1})$? That scaling makes statistical sense only when $E(Y_t/Y_{t-1}) = c$, but that is not generally the case for this particular data. Moreover, even if all the variables involved have constant mean, these can all be recovered from the constant term in the particular equation; see the parameterization of $\beta_0$ in table 1.

The obvious gap between the theoretical variables and the observed data is totally ignored! It is well-known that any equations that result from of individual optimization would denote intentions in light of a range of hypothetical choices. The data measure what actually happened, which would usually emerge from millions of interactions among agents. That is, what is observed often relates to the adjustment of the various quantities of interest as they emerge under ever changing market conditions, which is deliberately assumed away when equilibrium is imposed!

The DSGE structural model $M_\phi(z)$ is specified in terms of the observables: $Z_t = (Y_t, C_t, P_t, R_t)$.

This is derived by ‘solving’ a system of linear expectational difference equations that separates the observed from the latent variables, and then render it estimable with data $Z_0$. After eliminations, the transformed observable variables in terms of which the structural (estimable) model $M_\phi(x)$ is specified are:

$$y_t = \ln(Y_t/Y_{t-1}) - \ln(Y_t/Y_{t-1}), \quad p_t = \ln(P_t/P_{t-1}) - \ln(P_t/P_{t-1}), \quad r_t = \ln(R_t) - \ln(R_t),$$

Data: US quarterly time series for the period 1948-2010 ($n=252$):
$Y_t$ - per capita GDP, $P_t$ - consumers price index (CPI),
$r_t$ - gross interest rate on 90 days Treasury bill.

The statistical model $M_\theta(z)$, implicit in the DSGE structural model $M_\phi(z)$, is a Normal Vector AutoRegressive of order 2, [VAR(2)] model (table 3), popularized by Sims (1980), for $Z_t := (y_t, p_t, r_t)$, allowing $u_t$ to be autocorrelated. The Normal VAR(2) constitutes a parameterization of the process $\{Z_t, t\in\mathbb{N}\}$ assumed to be Normal, Markov(2) and Stationary.
The statistical ($\theta \in \Theta$) and the structural ($\varphi \in \Phi$) parameters are related via the implicit system of restrictions $G(\theta, \varphi) = 0$.

Table 3: Normal VAR(2) model

<table>
<thead>
<tr>
<th>Statistical GM: $Z_t = a_0 + A_1^T Z_{t-1} + A_2^T Z_{t-2} + u_t, t \in \mathbb{N},$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Normality: $D(Z_t, Z_{t-1}, \ldots Z_1; \theta)$ is Normal</td>
</tr>
<tr>
<td>[2] Linearity: $E(Z_t</td>
</tr>
<tr>
<td>[3] Homosked.: $Var(Z_t</td>
</tr>
<tr>
<td>[4] Markov: ${Z_t, t \in \mathbb{N}}$ is a Markov(2) process</td>
</tr>
<tr>
<td>[5] t-invariance: $\theta := (a_0, A_1, A_2, \Sigma)$ are t-invariant for all $t \in \mathbb{N}$</td>
</tr>
</tbody>
</table>

Estimation. The estimated parameters of the statistical model, a VAR(2) model are given in Table 4 with p-values in square brackets. The summary of results reported are described in detail in Poudyal and Spanos (2011).

Table 4: Estimation: Normal VAR(2)

<table>
<thead>
<tr>
<th></th>
<th>$y_t$</th>
<th>$p_t$</th>
<th>$r_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>.526[.000]</td>
<td>.041[.458]</td>
<td>.003[.893]</td>
</tr>
<tr>
<td>$y_{t-1}$</td>
<td>.305[.000]</td>
<td>.038[.185]</td>
<td>.025[.040]</td>
</tr>
<tr>
<td>$p_{t-1}$</td>
<td>.193[.191]</td>
<td>.549[.000]</td>
<td>.032[.232]</td>
</tr>
<tr>
<td>$r_{t-1}$</td>
<td>.015[.360]</td>
<td>.243[.117]</td>
<td>1.126[.000]</td>
</tr>
<tr>
<td>$y_{t-2}$</td>
<td>.096[.137]</td>
<td>.020[.476]</td>
<td>.012[.325]</td>
</tr>
<tr>
<td>$p_{t-2}$</td>
<td>-.243[.104]</td>
<td>.179[.006]</td>
<td>.035[.203]</td>
</tr>
<tr>
<td>$r_{t-2}$</td>
<td>-.197[.582]</td>
<td>-.114[.459]</td>
<td>-.191[.004]</td>
</tr>
</tbody>
</table>

Mis-Specification (M-S) testing. The M-S results reported in Table 5 reveal that the VAR(2) model is seriously misspecified, with departures from all assumptions but [2]. Therefore, no reliable inferences can be drawn on the basis the estimated VAR(2) model, however informal.

Table 5: M-S testing results: Normal VAR(2) model

<table>
<thead>
<tr>
<th>Model assumption</th>
<th>$y_t$</th>
<th>$p_t$</th>
<th>$r_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normality</td>
<td>.982[.008]</td>
<td>.901[.000]</td>
<td>.791[.000]</td>
</tr>
<tr>
<td>Linearity</td>
<td>1.44[.232]</td>
<td>0.607[.437]</td>
<td>.070[.792]</td>
</tr>
<tr>
<td>Homoskedasticity</td>
<td>5.299[.006]</td>
<td>37.285[.000]</td>
<td>3.401[.035]</td>
</tr>
<tr>
<td>Markov(2)</td>
<td>0.348[.706]</td>
<td>3.488[.032]</td>
<td>11.624[.000]</td>
</tr>
<tr>
<td>t-invariance</td>
<td>12.008[.000]</td>
<td>50.542[.000]</td>
<td>2.593[.077]</td>
</tr>
</tbody>
</table>

Respecification. The departures from Normality and Homoskedasticity, but not from Linearity suggest replacing the original Normality with a another distribution from the Elliptically Symmetric family; see Spanos (1994). Moreover, the rejection of the Markov(2) assumption suggests increasing the lag order. In light of this diagnosis, the process $\{Z_t, t \in \mathbb{N}\}$ is now assumed to be Student’s, Markov(3) and Stationary, giving rise to the Student’s VAR(3) model (Table 6).
Estimation of the Student’s VAR(3) model yielded the results in table 7, which, at first sight, differ significantly from those of the Normal VAR(2) in table 4.

<table>
<thead>
<tr>
<th>Table 7: Student’s t VAR(3)</th>
<th>Table 4: Normal VAR(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_t )</td>
<td>( p_t )</td>
</tr>
<tr>
<td>( \delta_0 )</td>
<td>.493[.000]</td>
</tr>
<tr>
<td>( t )</td>
<td>-.159[.000]</td>
</tr>
<tr>
<td>( t^2 )</td>
<td>1.060[.000]</td>
</tr>
<tr>
<td>( y_{t-1} )</td>
<td>.285[.000]</td>
</tr>
<tr>
<td>( p_{t-1} )</td>
<td>.287[.023] ( ^* )</td>
</tr>
<tr>
<td>( r_{t-1} )</td>
<td>-.607[.061] ( ^* )</td>
</tr>
<tr>
<td>( y_{t-2} )</td>
<td>.110[.027]</td>
</tr>
<tr>
<td>( r_{t-2} )</td>
<td>.273[.671] ( ^* )</td>
</tr>
<tr>
<td>( y_{t-3} )</td>
<td>-.205[.001]</td>
</tr>
<tr>
<td>( p_{t-3} )</td>
<td>-.446[.021]</td>
</tr>
<tr>
<td>( r_{t-3} )</td>
<td>.215[.606]</td>
</tr>
</tbody>
</table>

Moreover, the estimated skedastic function indicates highly significant coefficients, assuming that the estimated model is statistically adequate. A typical example is given below for \( y_t \):

\[
Var(y_t | Z_{t-1}^0) = 0.133 + 0.104\tilde{y}_t^2 + 0.773\tilde{p}_t^2 + 4.887\tilde{r}_t^2 + 0.112\tilde{y}_{t-1}^2 + 0.935\tilde{p}_{t-2}^2 + 12.283\tilde{r}_{t-2}^2 + 0.100\tilde{y}_{t-3}^2 + 0.783\tilde{p}_{t-3}^2 + 4.833\tilde{r}_{t-3}^2 + 0.065\tilde{p}_{t-1}\tilde{y}_{t-1} - 0.173\tilde{r}_{t-1}\tilde{y}_{t-1} - 0.027\tilde{y}_{t-1}\tilde{p}_{t-2} \cdots
\]

where \( \tilde{x}_t := (x_t - \overline{x}) \).

Thorough M-S testing of the estimated Student’s VAR(3) model indicated no departures from its assumptions; see table 8. Its statistical adequacy secures the reliability of any inferences based on it, including the data-pertinacy of the DSGE model in table 2.
Table 8: M-S testing results: Student’s t, VAR(3) model

<table>
<thead>
<tr>
<th>Model assumption</th>
<th>$y_t$</th>
<th>$p_t$</th>
<th>$r_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s t</td>
<td>1.919[0.383]</td>
<td>2.056[0.358]</td>
<td>4.401[0.111]</td>
</tr>
<tr>
<td>Linearity</td>
<td>2.838[0.096]</td>
<td>0.564[0.458]</td>
<td>2.422[0.123]</td>
</tr>
<tr>
<td>Heteroskedasticity</td>
<td>0.017[0.733]</td>
<td>1.406[0.426]</td>
<td>4.741[0.104]</td>
</tr>
<tr>
<td>Independence</td>
<td>0.312[0.729]</td>
<td>0.448[0.602]</td>
<td>1.128[0.251]</td>
</tr>
<tr>
<td>t-invariance</td>
<td>0.976[0.375]</td>
<td>0.853[0.325]</td>
<td>0.001[0.982]</td>
</tr>
</tbody>
</table>

The sizeable discrepancies [indicated by [.]•] between the coefficient estimates of the Student’s t VAR(3) and the Normal VAR(2) give the first indication that the structural DSGE is data-impertinent. This is formally confirmed by confronting the DSGE model with the data $Z_0$ via testing its over-identifying restrictions:

$$H_0: \mathbf{G}(\theta, \varphi)=0, \ vs. \ H_1: \mathbf{G}(\theta, \varphi)\neq 0,$$

for $\theta \in \Theta$, $\varphi \in \Phi$, in the context of the statistically adequate Student’s t VAR(3) model. The relevant test is based on the likelihood ratio statistic:

$$\lambda_n(Z) = \frac{\max_{\varphi \in \Phi} L(\varphi; X)}{\max_{\theta \in \Theta} L(\theta; X)} \Rightarrow -2 \ln \lambda_n(X) \overset{H_0}{\sim} \chi^2(m). \quad (14)$$

For $m=27$, for $\alpha = .05$, $c_{a} = 40.1$, the observed test statistic yields:

$$-2 \ln \lambda_n(x_0) = 5662.13[.000000000].$$

This result provides indisputably strong evidence against the DSGE model!

What is interesting about the DSGE modeling is that all the potential modeling and inference mistakes anticipated by Haavelmo (1944) are committed routinely. It is well-known that a theoretical model that stems from any form of optimization would denote intentions (schedules) in light of a range of hypothetical choices. The data measure what actually happened, which would usually emerge from millions of interactions among agents. Hence, what is observed often relates to the adjustment of the realized quantities of interest as they emerge under ever changing market conditions. That is, what is observed is exactly what is assumed away by DSGE modeling when equilibrium is imposed!

Haavelmo (1989) reminded econometricians about the gap between theory and data in his Nobel lecture:

"It is quite obvious that if the theories we build to simulate actual economic life are not sufficiently realistic, that is, if the data we get to work on in practice are not produced the way that economic theories suggest, then it is rather meaningless to confront actual observations with relations that describe something else." (Haavelmo, 1989, p. 14)

"I think it is not unfair to describe major part of existing economic theory in the following way. We start by studying the behavior of the individual under various conditions of choice. Some of these conditions are due to the fact that the individual has to have contact with in his economic affairs with other individuals. We then try to construct a model of the economic society in its totality by so-called process of aggregation. I now thing this is actually beginning at the wrong end. Speaking very briefly and along very
broad lines, I think that economic theory could make progress by an approach within the following framework.” (Haavelmo, 1989, p. 15)

Another major problem with the DSGE models is that no attempt is made to account for the probabilistic structure of the data, and the statistical adequacy of \( M_\theta(z) \) implicitly assumed by their structural model \( M_\varphi(z) \) is completely ignored. As a result, these models enjoy only a perilous link to the macro-phenomena they aim to shed light on (describe, explain predict) and give rise to totally untrustworthy evidence; see Poudyal and Spanos (2011). Indeed, the DSGE models display an ostensible degree of ‘theoretical rigor’ that is inversely proportional to their pertinence in modeling observable macroeconomic phenomena.

Summary report for DSGE modeling

a. The DSGE models aim to explain key macroeconomic phenomena like business cycles and growth in terms of individual intertemporal decision making facing stochastic shocks. The so-called rigorous microfoundations constitute a misplaced justification for the Ricardian vice. The theory-oriented transformations of the data are totally misplaced and statistically uninformed.

b. The widely held belief that DSGE modeling provides natural links to data via simulated trajectories of the variables involved, correlations among observables and impulse respond analysis, on closer examination constitutes a vice, not a virtue, because such links are anemic at best and lack credibility without statistical adequacy!

c. Using the Bayesian approach to statistically analyze DSGE models adds significantly to the untrustworthiness of the resulting evidence by glossing over the statistical misspecifications. Bayesian inference is equally vulnerable to statistical misspecification because a misspecified \( M_\theta(z) \) gives rise to an invalid likelihood \( L(\theta; z_0) \propto f(z_0; \theta) \) and thus an erroneous posterior \( \pi(\theta \mid z_0) = \pi(\theta) \cdot L(\theta; z_0) \), for all \( \theta \in \Theta \).

d. Claims like “it is shown that the estimated model is able to compete with more standard, unrestricted time series models, such as vector auto regressions (VARs), in out-of-sample forecasting” are simply false, unless the comparison models are equally misspecified! The out-of-sample forecasting performance of the above estimated DSGE model is terrible compared to the unrestricted Student’s t VAR(3) model!

8 Conclusions

The Haavelmo (1944) monograph constitutes the best example of viewing theory testing in the context of bridging of the gap between theory and data, where both the theory and the data are accorded ‘a life of their own’. It contains a wealth of methodological insights, which, unfortunately for econometrics, had no impact on the subsequent developments of textbook econometrics.

One of the main objectives of this paper is to bring out some of Haavelmo’s crucial methodological ideas and insights with a view to argue, first, that their neglect contributed significantly to the endemic untrustworthiness of empirical evidence in
economics, and second, to make a case that one can integrate some of these key ideas and insights into a refined/extended Fisher-Neyman-Pearson statistical approach to inference, to construct a more pertinent modeling framework for empirical modeling in economics. The key to this blending is the distinction between substantive and statistical adequacy, which is used to shed light on several confusions permeating textbook econometrics. This error statistical framework can ensure learning from data by securing the reliability of inference and thus enhancing the truthworthiness of empirical evidence.

The error statistical perspective is used to unveil the untrustworthiness of empirical evidence in the case of the DSGE modeling that dominates current macroeconometrics, as well as illustrate the perspicacity and current relevance of some of Haavelmo’s disregarded ideas. The discussion also makes it clear that "there is no Royal Road to econometric modeling, and really valuable ideas and insights can only be had at the price of good grounding in probability and hard work":

“We believe that, if economics is to establish itself as a reputable quantitative science, many economics will have to revise their ideas as to the level of statistical theory and technique and the amount of tedious work that will be required, even for modest projects of research. On the other side we must count the time and work that might be saved by eliminating a good deal of planless and futile juggling with figures.” (Haavelmo, 1944, p. 114)

References


28


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.


