

## **Lessons in Econometric Methodology: The Axiom of Correct Specification**

**Asad Zaman<sup>®</sup>**

International Islamic University of Islamabad

### **ABSTRACT**

Leamer first pointed out that a regression model is valid only if all of the assumptions under which it is constructed are valid. In particular, this means that all of relevant regressors which are determinants must be included. In practice, applied econometricians assume that whatever model they put is valid, when there can only be one valid model. If any relevant regressor is omitted, then the equation is mis-specified and conclusions drawn from the regression can be seriously misleading. Even though this mis-specification analysis is included in textbooks, it is routinely ignored in applications, where researcher interpret their equations as if they have correctly specified all regressors, without testing to see if this may be the cases. Hendry's encompassing methodology provides a remedy for this problem, but seems to be unfamiliar to many. The purpose of this pedagogical note is to provide an introduction to an elementary but important aspect of this methodology.

**Key words:** *Econometric Methodology, Regression Model, Specification, Missing Variable, Model Misspecification*

JEL Classifications: C18, C51

### **1. INTRODUCTION**

Based on my experiences as an Editor & Referee for papers submitted to several journals, it is clear that large numbers of people have major misconceptions about econometric methodology. Furthermore, this is not their fault; standard econometric textbooks typically do not address these methodological problems, so that students all over the world are trained to make these mistakes. The discussion is structured around Leamer's (1978) criticism of textbook inference in regression models. He writes that the "textbook version" of inference in regression model can easily be dismissed, because it is based on the unacceptable "Axiom of Correct Specification". This axiom is the following set of assumptions, which are not explicitly mentioned in textbooks, but nonetheless required for inference:

- 1) The set of explanatory variables that are thought to determine (linearly) the dependent variable must be
  - a) unique,
  - b) complete,
  - c) small in number, and
  - d) observable.
- 2) Other determinants of the dependent variable must have a probability distribution with at most a few unknown parameters.
- 3) All unknown parameters must be constant.

---

<sup>®</sup> Asad Zaman, International Institute of Islamic Economic, International Islamic University of Islamabad, (email: [asadzaman@alum.mit.edu](mailto:asadzaman@alum.mit.edu)), Tel. +9251 9257939, Fax: + 9251 9258019

The methodological mistakes being taught in conventional textbooks have not changed since then. Hendry (1995) discusses Leamer's Axiom and renames it the Axiom of Omniscience: Only someone who know exactly what the true model is can run a valid regression; all others have no chance of doing so. Although econometrics textbooks do discuss the "assumptions" of the regression model, they proceed to completely ignore the significance and importance of these assumptions. In particular, they fail to point out that if these assumptions are violated, the regression results will be invalid. If these assumptions are taken seriously, then nearly all regressions being run today are invalid, or 'nonsense regressions'. As a result, David Freedman (2010) wrote that "An enormous amount of fiction has been produced, masquerading as rigorous science".

As famous philosopher of science Karl Popper realized, scientific theories can never be proven to be valid, but they can be proven to be false. The same is true of econometric models. We can never prove them to be "true," but we can prove them to be false. Conventional methodology is based on assuming that the model we specify is true, and never testing this assumption. This can be called the "ostrich" methodology, since even obvious flaws and defects are overlooked. The main skill which applied econometricians need to learn is how to prove regression models to be false. In this pedagogical note, we will focus on only one method for achieving this goal, which is often overlooked, and mis-understood, based only on the requirement (1) specified in the Axiom of Correct Specification. A regression model must correctly specify the unique small set of observable regressors which determine the dependent variable. This implies that a regression model specifies the variables which are included in the regression model, and also, at the same time, the variables which are excluded from the regression model. These exclusions are just as important as the inclusions. In particular, if an important variable is left out of the model, the model is mis-specified, and the omission can lead to seriously biased and misleading estimates. A key element in establishing validity of a regression is to ensure that no important relevant regressors have been omitted from it. Because this elementary point is routinely ignored in practical applications of econometrics, we explain it in detail below.

## **2. A TYPICAL APPLIED ECONOMETRICS PAPER**

I will discuss three actual papers which were submitted, evaluated, and rejected, much to the dismay and distress of their authors, who considered them to be very good papers. Indeed, these papers reflected best practices in conventional econometrics in the 1970's, before the publication of Davidson, Hendry, Srba and Yeo (1978). This paper pointed out serious defects in conventional econometric methodology, and introduced a new approach, which I will call the DHSY methodology. This has also been called the Hendry methodology, as well as the London School of Economics (LSE) methodology in the literature; see Faust and Whiteman (1997) for a discussion and further references. I will focus on only one element of this methodology, namely General-To-Simple (GeTS) modeling which is the opposite of the conventional econometric practice of simple-to-general modeling. Before explaining the reasons for this reversal of conventional practice, I would like to summarize some elements of the conventional methodology, as encapsulated in the three papers under discussion.

The general pattern of these papers is as follows. The authors want to investigate the effect of  $X$  on  $Y$ ;  $Y$  is generally taken to be economic growth, since that is considered to be the most important variable. The authors do a literature review on the effects of  $X$  on  $Y$ . From this

literature review, they identify additional relevant variables  $Z_1, Z_2, \dots, Z_r$  and estimate a regression equation of the type

$$Y = a + bX + c_1Z_1 + \dots + c_rZ_r + \varepsilon$$

They note that the  $R$ -squared is reasonable, the signs are correct, and most variables are significant, and proceed to interpret the results. They use the signs of the estimated coefficients to assess whether the independent variables impact on  $Y$  positively and negatively, and the significance to assess whether they are “determinants” of  $Y$  or not.

This conventional methodology, commonly used by aspiring econometricians all over the world, is deeply flawed, as noted by Leamer, and many others. Unfortunately, the dire consequences of the Axiom of Correct Specification, or Omniscience, have not been widely absorbed. As a result, authors of papers following this conventional methodology have difficulty understanding why their papers are rejected, and sometimes attribute it to personal biases of editors.

### 2.1. Model A: GDP on FDI and Human Capital

After doing some preliminary analysis, model A estimates the following equation, which related  $\log(GDP)$  to  $\log(FDI)$  and  $Lit(eracy)$  taken as a proxy for human capital. In addition,  $\log(EXPorts)$  and an interaction term between  $FDI$  and human capital are added. After putting in this fairly ad-hoc set of variables on the right hand side, authors of paper A estimate a regression, find that all the coefficients are significant, and proceed to interpret the meaning of this regression and derive policy implications for Pakistan. Their estimated equation is as follows:

$$\log(GDP) = 10.93 + 3.65 \log(FDI) + 0.076 Lit - 3.65 \log(FDI \times Lit) + 0.64 \log(EXP) + \varepsilon$$

(1.42)    (0.53)                    (0.012)    (0.53)                    (0.07)                    ( $R^2 = 0.95$ )

The authors find positive coefficients for  $FDI$  and  $Literacy$ , and conclude that these variables have a positive effect on  $GNP$ , but the interaction term has a negative effect. Similarly,  $\log(EXPorts)$  are highly significant with a coefficient of 0.6, so that every percentage increase in exports adds 0.6 percent to the  $GDP$ . Are any of these conclusions justified? Would we feel comfortable in taking these results and basing policy on them? Can we conclude that a 1% increase in  $FDI$  will generate 3.65% increase in  $GNP$ , as the estimated regression equation suggests? Can we conclude that if both  $FDI$  and  $Literacy$  in Pakistan increase at the same time, this would have a negative impact on the  $GDP$  through the interaction term?

Before we discuss this further, let us turn to another paper, which does a very different analysis of the  $GDP$ :

### 2.2. Model B: GDP and Intellectual Property Rights

Here the authors are concerned with the impact of Intellectual Property Rights ( $IPR$ ) on  $GDP$ . They provide a very nice literature review which show how  $IPR$  can impact on growth. As additional variables, authors take into consideration  $FDI$ ,  $EFW$  (economic freedom index), trade openness, population growth, secondary years of education ( $SYR15$ ), and  $GDI$ , gross domestic investment. They estimate a long run equation of the following form:

$$GDP = 121.76 IPR + 0.06 FDI + 33.04 EFW + 0.01 TRADEOP + 131.90 PopGrowth + 22.15 SYR15 + 0.004 GDI$$

Their main conclusion is the Intellectual Property Rights are highly significant as a determinant of *GDP* with a high impact coefficient of 121.76. From this equation they can derive policy implications that since all of the variables have positive and significant coefficients; Pakistan can prosper by imposing stringent policies protecting Intellectual Property Rights.

As a final illustration, we consider paper C, which examines the export led growth hypothesis.

### **2.3. Model C: Export Led Growth**

Model C is the simplest of all. It regresses the log of gross *GNP* (NY.GDP.MKTP.CN) of the log of gross exports of Pakistan (NE.EXP.GNFS.KN) from the WDI data set, getting the following regression:

$$\log(GDP) = -15.47 + 2.57 \log(EXP) + \varepsilon; R^2 = 0.96$$

The author finds strong support for the export led growth hypothesis. The elasticity of *GDP* with respect to exports is 2.57, which means that a 1% increase in exports will lead to a 2.57% increase in the *GDP*. Since *GDP* has been growing at less than 5% recently, this equation shows us a quick and easy way to get fantastic rates of *GDP* growth. We only need to focus on increasing our exports.

### **2.4. The Million Dollar Question**

The Million Dollar question is: Is it possible that all three authors are correct? That is, paper A is a proper analysis about the impact of *FDI* on *GDP*, while paper B provides us with valid information about the impact of Intellectual Property Rights, and paper C tells us about the impact of Exports on *GDP*? I was surprised to find that many students to whom I posed this question said “YES” – each of the three authors has followed correct methodology to arrive at a solution to the problem that they were investigating. In fact, we can be quite sure that all three models are seriously wrong, and using them to try to interpret and understand economic relationships in Pakistan would be a serious mistake. We will now explain how we can be sure that these are all nonsense regressions.

## **3. IMPLICATIONS OF THE SPECIFICATION AXIOM**

Everybody who takes regression analysis course, studies the assumptions of regression model. But nobody knows why, because after reading about the axioms, they are rarely mentioned. But the assumptions are important, because if any one assumption is wrong, the regression is not valid, and the interpretations can be completely wrong. In order to have a valid regression model, you must have right regressors, the right functional form, all the regressors must be exogenous, regression parameters should not change over time, regression residuals should be independent and have mean zero, and many other things as well. There are so many assumptions that it is impossible to test all of them. This means that interpreting a regression model is always a matter of FAITH – we must BELIEVE, without having any empirical evidence, that our model is the ONE TRUE VALID model. It is only under this assumption that our interpretations of regression models are valid. That is why Hendry has called it the Axiom of Omniscience.

Although the Axiom of Correct Specification requires much more, we will focus in this paper on only one implication. This is that we must specify all the regressors correctly, using the

correct functional form, in order to have a valid regression. Unless we start with the perfect model, our results will be wrong:

**Nonsense Regressions:** If a regression model **OMITS** a significant regressor then it is **INVALID**; we may call such regressions “nonsense regressions”.

This formulation highlights the major mistake in modeling that is common. The regressors which are **EXCLUDED** by a regression model are just as important as the ones that are included. Thus the simple model C not only states that FDI determines GDP, it also states that no other variable has any effect on GDP, since no other variable is included in the model. It is this exclusion which is seriously questionable. Now, we come back to our Million Dollar question – is it possible for all three regression models A, B, and C, to be correct?

### 3.1. The Million Dollar Question

By now it should be obvious that these three models contradict each other. This is because the FDI model says that IPR does not belong in the regression equation. Consider adding IPR to the FDI equation:

$$\log GDP_t = a + b_1 \log FDI_t + b_2 Lit_t + b_3 \log FDI \times Lit_t + b_4 \log Exp_t + b_5 IPR_t + \varepsilon_t$$

If model A is correct then  $b_5 = 0$ . If this coefficient is not zero, then the equation is missing a variable which should be in the regression. In this case the equation is mis-specified and suffers from omitted variable bias. All of the estimates of the other terms are biased in this case. So, if model A is a correct model, then models B and C **MUST** be wrong. This is because model B says that *IPR* a significant determinant of *GDP* but model A says that *IPR* does not matter for determining *GDP*. Similarly, model A says that *FDI* and *Lit* are significant, while model C says that these variables do not matter, so both cannot be true at the same time.

Every regressor excluded from a regression model is automatically included in the error term of the regression model. Students will recall that independence of the errors and the regressors is one of the crucial assumptions of regression. If the excluded regressors are independent of the included regressors, then the exclusion will not cause bias in the estimated coefficients. However, if the excluded regressors are significant, and are correlated with the included regressors, this will create serious bias in the estimates.

So, we come to the conclusion that it is impossible for **ALL** three models to be correct at the same time. At least two, and maybe all three, models are invalid, and are examples of nonsense regressions. Making policy on the basis of these regressions would lead to dangerously poor decisions. Next, is there any reason to prefer any one of the models to the others? Is A superior to B and C for example? Students often think that this matter can be decided on the basis of *R*-squared, but this is a wrong training imparted by conventional textbooks. *R*-squared is useless for choosing among models, and even the adjusted *R*-squared is not of any help. Other fancy model selection criteria like AIC, BIC, Schwartz, etc. are of no use in discriminating between valid and invalid regressions. Completely wrong models can have very good fit, as well as high *R*-squared values, as we will show.

The fundamental problem is the methodological mindset created by conventional textbooks that any regression model I write down is automatically valid. Once I estimate a regression model, and it passes basic checks, then I am authorized to interpret it, and derive policy

implications. This is exactly the opposite of the mindset required by the Axiom of Correct Specification. The chances of my identifying exactly the right set of regressors, in exactly the correct functional form, are virtually zero. Therefore, almost any regression I estimate will be spurious, or invalid. Therefore, the job of the econometrician BEGINS after the estimation, and this job requires the production of evidence that the estimated regression is valid. Regression models can only provide some hints and clues about reality, and we must follow them up, and provide supporting evidence from many different sources to support the regression results. There are many ways that we can approach this problem. The simplest is to follow Popper's scientific methodology. We can TRY to prove that the regression equation we have estimated is wrong. If we try very hard, but fail to do so, then there is a possibility that the estimated equation may be right. This is the basis of the DHSY methodology, also known as the Hendry Methodology, or the LSE Methodology, developed and demonstrated in a justly famous paper by Davidson et al. (1978). For a clear exposition of the details of the Hendry Methodology, and how it is superior to conventional methodology, see Gilbert (1986). The essence of the Hendry methodology can be summarized in three principles: "Test, Test, and Test". That is, if our final equation passes all possible tests for mis-specification which we can devise, then it has a chance of being the correct equation. We consider this further below.

### 3.2. How Much Difference Can a Missing Regressor Make?

What students (and sometimes even professors) do not realize is that if a major regressor is missing from a regression equation, then all bets are off – any kind of result can occur. To illustrate this, consider first a standard consumption equation for Pakistan, where we regress the consumption of Pakistan on the *GDP* of Pakistan, taken from the WDI data set:

*CONS* = Final consumption expenditure, etc. (constant 2010 US\$) = NE.CON.TETC.KD  
*GDP* = GDP (constant 2010 US\$) = NY.GDP.MKTP.KD

$$\begin{array}{rcccl} \text{CONS} = 4.12 + 0.883 \text{ GDP} + \varepsilon & & (R^2 = 0.998) \\ (0.51) (0.006) & & (2.56) \end{array}$$

The low standard errors, written in parentheses, show that estimated coefficients are highly significant. In this situation, where we know in advance from theoretical and empirical evidence (very high  $R^2$ ) that Income is the major and principal determinant of consumption; the Axiom of Correct Specification does not create massive difficulties for us. Although there are sure to be missing variables, the unexplained variation in consumption is sufficiently small that biases introduced by omitted variables should be small. Another way to see this is to note that *CONS* and *GDP* are co-integrated, and in such situations, the phenomenon of "super-consistency" obtains; see Stock (1987) for theoretical details. In particular, this implies that the OLS estimates will converge rapidly to the true value even if there are missing stationary regressors, or if short term dynamics are mis-specified in the equation.

However, the situation changes completely if major determinants of the dependent variable are omitted from the equation. As an illustration, consider regression the same variable *CONS* on two other variables available in the WDI Data Set on Pakistan:

*SUR* = Survival to age 65, female (% of cohort) = SP.DYN.TO65.FE.ZS

*CO2* = CO2 emissions from gaseous fuel consumption (% of total) = EN.ATM.CO2E.GF.ZS

Obviously, these variables have no relation to Consumption. Nonetheless, the OLS regression yields the following results:

$$\text{CONS} = -268.7 + 6.78 \text{ SUR} - 1.82 \text{ CO}_2 + \varepsilon \quad (R^2 = 0.84)$$

(25.9) (0.73) (0.65) (20.0)

Both variables are highly significant, with  $p$ -values well below 1%. So, based on this regression, we could claim that the major determinants of Pakistani Consumption are the Carbon Dioxide Emissions, and the percentage of females who survive to the age of 65. These two variables together determine about 84% of the variation in Consumption, so that any remaining determinants can only account for at most 16% of the unexplained variation in consumption. Of course, this is nonsense, and the above regression is a nonsense regression, just like models A, B, and C for Pakistan *GNP* are nonsense regressions. In this particular case, it is easy to show that the above equation is nonsense by simply including the relevant regressor in the equation. This leads to:

$$\text{CONS} = 15.64 + 0.902 \text{ GDP} - 2.60 \text{ SUR} + 67.8 \text{ CO}_2 + \varepsilon \quad (R^2 = 0.99)$$

(0.54) (0.014) (1.40) (80.3) (2.28)

In this equation, the estimated coefficients of SUR and CO<sub>2</sub> have high standard errors and are not significant at 5% level. This shows that they became significant only because of the omitted variable GDP. Once the principal determinant, or a group of important variables, is omitted from the regression equation, then any other variables can and will become significant if it is correlated with the omitted variables. If we omit GDP from the above equation for Pakistani consumption, and take any other variable from the WDI data set which has an increasing trend, we will get a significant regression. For example, we could take the GNP of Honduras and get the following highly significant regression equation.

$$\text{CONS(Pakistan)} = -13.44 + 11.07 \text{ GDP(Honduras)} + \varepsilon \quad (R^2 = 0.99)$$

(1.15) (0.12) (4.34)

From this equation, we see that the GDP of Honduras is a very powerful and significant determinant of the Consumption in Pakistan. This is obviously nonsense, but there is nothing in the regression which tell us so. In fact, it is the variable which is missing from the regression, namely the GDP of Pakistan, which can tell us that this is regression is wrong and misleading.

### 3.3. Implications of Omitted Variables

As we have seen, if a significant regressor is omitted from a regression equation, the results can be completely wrong and misleading. But how can we tell whether or not a significant variable has been omitted? The regression itself will not provide us with any clues. All of the statistics can be very healthy, with high  $R$ -squares, significant  $t$ -statistics for all regressors, the correct signs, and everything else one could hope for in a regression. The missing variable does not signal its absence by any observable defects in the estimated equation.

There is one situation where problems posed by the Axiom of Correct Specification have a potential solution. This is when theory and empirical evidence inform us that there is a very small set of regressors which determine the dependent variable. The Keynesian Consumption function is one example of this situation. Here, we believe that current consumption is dependent on current income, savings, wealth and habit-formation. This limited set of variables makes sense intuitively, matches our subjective and introspective evidence, corresponds to what we learn from consumer surveys, and is also affirmed by huge numbers of empirical studies of the consumption function. Because of all of this evidence, if we are

careful, we can be reasonably sure of taking into account all major determinants of consumption in regression equations. There remain questions of choosing the right functional form, the right definitions of the variables in question, and other nitty gritty details. In fact, drastic variations of results in precisely this situation is what led to the development of the Hendry Methodology, which we will discuss in the next section. However, the problem seems manageable in this context.

The other situation, where theory offers no guidance, and potentially any variable is a determinant, is rather more difficult. For example, suppose we only have ten possible determinants for our independent variable. Then there are  $2^{10} = 1024$  possible linear regression models, out of which only one is correct. And this ignores the possibility of nonlinear functions like squares, cubes, logs and other transforms of the independent variable. On top of this, we have the possibility of interaction terms which would quickly lead to millions of possible regressions, out of which only one can be correct. Here too the Hendry methodology offers a potential solution, but the chance of our going astray and picking the wrong set of regressors seem much higher in this case. We will discuss how these situations can be handled in a later paper. It seems likely that tools of exploratory data analysis will be much more suitable than those of confirmatory data analysis. These concepts were introduced by Tukey (1977) and provide a lot of insight into two different modes of data analysis. Since regression textbooks deal principally and mainly with confirmatory data analysis, these methods will not be of much use in these situations.

#### 4. POTENTIAL SOLUTIONS

As we have discussed, omitted variables pose a significant threat to regression models, and imply that the original regression could be a purely nonsensical regression. A natural way to try to solve this problem would be to add potentially relevant regressors  $W_1, \dots, W_m$  to the set of regressors  $X_1, \dots, X_k$  into a regression model for  $Y$ . If at least one of the regressors  $W$  proves significant, then it is immediately obvious that the original model is missing significant regressors. So far, this reasoning is correct, but the natural remedy of adding  $W$  to fix the problem is not correct. That is because the significance of the added variable  $W$  does not imply that  $W$  is the missing variable. As we have seen, once an important variable is missing, any other variable which is correlated with the missing variable will appear to be significant. For instance, when Pakistan *GDP* is omitted from the equation for Pakistan Consumption, then the Carbon Dioxide Emissions, the Percentage of Females who survive to age 65, and the *GDP* of Honduras, all become significant explanatory variables for the Consumption. If variable  $W$  is significant when added to a regression model, we cannot conclude that  $W$  is the missing variable. We can only conclude that there is some variable which has been omitted, and  $W$  is correlated with the missing variable.

The fact that inferences can change dramatically if we change the sets of regressors was noted by Leamer (1978), and labelled the ‘fragility’ of conventional inference. The methodological theory taught in textbooks demands that a well specified model should exist in advance of empirical estimation; that is, we must know the true model, before we run the regression. The reality is very different. In practice, applied econometricians search through hundreds of models, looking for coefficients which match their presuppositions. As Huff (1954) noted, “if you torture the data long enough, it will confess to anything.” As we have seen in the regressions above, the estimated coefficient of log Exports is 2.57 in model (A), and 0.64 in model (C), depending on which other regressors are included. In fact, if we add the level of total Industrial production to model (C) as a determinant of the GNP, we find that the



coefficient of Exports becomes negative and insignificant. So depending on which theory we have in mind, we can create a regression to support our theory by searching through the potential regressors, can create a story to justify our choice afterwards. For the relationship between exports and growth, Zaman (2010) show that we can find published papers which support all of the four possibilities: (1) Exports cause Growth; (2) Growth Causes Exports; (3) Bidirectional Causality; (4) No Causal Relationship.

#### **4.1. Extreme Bounds Analysis**

Leamer proposed extreme bounds analysis (EBA) as a way of dealing with fragility of inference. This method was originally proposed using a Bayesian justification, but it also has a simple and direct, intuitively appealing interpretation. Suppose that we want to examine a regression of  $Y$  on selected regressors  $X_1, \dots, X_k$  which are known to be determinants of  $Y$ . In addition, we have variables  $W_1, \dots, W_r$  about which we have some doubt as to whether or not they should be included. EBA proposes to try out all collections of variables and examine how the estimated coefficients change. If variable  $W$  remains significant regardless of how many combinations of doubtful variables we include or exclude, then inference about  $W$  is considered robust. If the coefficient of  $W$  changes substantially, and covers the range from positive to negative, then 0 is a possible value of the coefficient. Thus  $W$  may or may not belong to the regression equation, depending on which variables are included or excluded. Thus Leamer provides us with a procedure to assess whether or not the significance of variable  $W$  depends on variables which have been omitted from the equation. This responds to the difficult raised by the Axiom of Correct Specification that we have discussed in this article.

Leamer's original methodology, as described above, leads to pessimistic conclusions. First, if we have only a small set of observed regressors, there is a significant possibility that important and relevant variables have not been measured and included in the data set. So the method is worth applying only on rich data sets which have large collections of variables. Second, we find that in such situations, we usually cannot find any significant determinants. As we vary combinations of regressors, almost any estimate can arise on a sufficiently rich data set. Thus the conclusion is that inference is always fragile, and we can never learn anything about determinants from the data.

Many authors have criticized Extreme Bounds Analysis as being too extreme. Granger and Uhlig (1990) show that EBA involves trying out all regressions, including those with extremely poor  $R$ -squares, and hence having very bad fit to the data. They suggest modifying the method to try only those regressions which have reasonable high values of  $R$ -squares. A similar critique is made on Bayesian grounds in Zaman (1996, Chapter 16, Section 1, Remark 5, page 424). Leamer's procedure can be seen as using prior specifications which are highly implausible in light of the observed data. Using these priors is equivalent to safeguarding against extremely unlikely possibilities. For example, it can happen that taking into account the possibility that the  $GNP$  of Honduras is a significant determinant of Pakistani Consumption makes the coefficient of Financial Assets in Pakistan insignificant. Then Leamer's EBA would conclude that we cannot say whether or not ownership of Financial Assets matters for Pakistani Consumption. However, it would be more reasonable to rule out the  $GNP$  of Honduras as a relevant variable and to assume that Financial Asset ownership matters for consumption, even if it become insignificant when  $GNP$  of Honduras is included in the regression model. This is because theoretical considerations can over-ride empirical considerations, as will be discussed in greater detail in a later section of this paper.

## 4.2. I Ran Two Million Regressions

Leamer's EBA is based on the simple idea that we can check whether or not omitted variables can make a significant difference just by putting them into the regression equation and seeing if they matter. Unfortunately, the original EBA technique does not lead to useful results, generally discarding all variables as being insignificant. Some reasons for this problem have been explained in the previous section. Sala-i-Martin (1997) modifies the EBA technique slightly, to produce a version which does produce positive results. Instead of asking for a regressor to remain significant in combination with all possible omitted regressors, he requires that the regressor should be significant in 95% of all such regressions. This modification of EBA is illustrated in his paper entitled "I just ran two million regressions". First, the author examines the growth literature and identifies 62 variables which have been identified as being significant as determinants of growth (*GR*). Running all the possible  $2^{62}$  regressions lay beyond computational possibilities available to the author, so he used a simplifying strategy. He selected three variables which were considered significant universally, and included them in all of the regressions. These variables were *GNP* in 1960 (*GNP60*), Life Expectancy in 1960 (*LE60*), and Primary School Enrolment in 1960 (*PSE60*). These were considered as baseline measure of Wealth, Health, and Human Capital, widely accepted as relevant and important by researchers of growth. That leaves 59 regressors (call them  $X_1, X_2, \dots, X_{59}$ ) which are to be tested for significance. Sala-i-Martin picks one of these regressors – call it  $W$  – and tests it for significance as follows. As we have seen earlier, just putting in  $W$  in a regression of *GR* on *GNP60*, *LE60*, and *PSE60*, is not enough.  $W$  could be significant just because it is correlated with the real determinants of *GR* which are omitted from the equation.

Ideally, Sala-i-Martin wants to put in all possible combinations of the remaining variables, and assess whether  $W$  remains a significant variable in all such regressions. However, this task is too difficult computationally – the author estimates that it would require trillions of regressions, and take decades of computer time to complete. So, he simplifies that task by adding the remaining 58 regressors (after taking out  $W$  and the three fixed regressors) in groups of 3. This means running  $58 \times 57 \times 56 = 185,136$  regressions of *GR* on *GNP60*, *LE60*, *PSE60*,  $W$  and three regressors  $X(i), X(j), X(k)$ , where  $i, j, k$  are chosen from the remaining 58 regressors. He then considers  $W$  to be significant if it comes out as a significant variable in 95% of these 185,136 regressions. By including all possible groups of three, we ensure that  $W$  is compared with all possible groups of omitted variables, and selected as significant only if it is significant in combination with 95% of such groups.

As a result of this strategy, Sala-i-Martin finds that 22 variables out of the 58 considered qualify as significant determinants of growth. He groups these variables into categories which reflect political, religious, as well as domestic and international economic variables. Many criticisms can be made of this approach to establishing the significance of variables. However, Sala-i-Martin has dealt with the basic issue which is being highlighted in this note on methodology – a variable may appear to be significant only because it is correlated with an important omitted variable. The 22 variables selected as significant determinants of growth all have the property that they remain significant in combination with any three out of 58 other regressors. This means that large numbers of omitted variables have been tested for, so that we have some assurance that  $W$  is significant on its own merits, and not because it is acting as a proxy for some other variables. A useful computer package which allows the implementation of many sophisticated variants of the Extreme Bounds Analysis, together with a good discussion of the associated literature, is available in Hlavac (2016).

The conclusions of Sala-i-Martin do have an important implication. If their results are valid then there are 22 significant determinants of growth among the 62 variables considered. This means that any regression equation which excludes any of these 22 variables is seriously misspecified. If we put in any combination of variables – completely unrelated to growth – the coefficients estimated and their significance will depend entirely on how well these variables act as proxies for the omitted 22 variables. Any policy implications drawn from such regression will be completely wrong and misleading. Since I know of no growth regression which uses all 22 variables, it would seem that all the growth regressions reported in the literature are just nonsense regressions. Also, the Sala-i-Martin results cast doubt on his own methodology. Legitimate growth regressions must include all 22 significant regressors, to avoid omitted variable bias. However, all of the two million regressions run by Sala-i-Martin include only 7 regressors, and hence all of them are invalid (nonsense) regressions. It is not clear that we can get sensible results by running two million nonsensical regressions and analyzing their output. Another critique of the Sala-i-Martin methodology is presented in Section 5 of this paper.

### **4.3. A Bayesian Approach**

A purely Bayesian approach to the problem of finding the right model is advocated by Fernandez, Ley and Steel (2001). They re-analyze the results of Sala-i-Martin (1997) but restrict the analysis to 41 regressors for which complete data is available for all countries over the period chosen for their analysis. They face the problem of putting down Bayesian priors over all of the 241 models possible by selection from the 41 regressors available. They suggest some reasonable priors for which calculations are possible. Even though there are more than 2 trillion models involved, newly developed Markov Chain Monte Carlo techniques can use simulations to estimate the relevant posterior probabilities. Even though the technical details are quite gory, the intuition is very simple. A priori, each regressor is assigned a 50% probability of belonging to the model. After assigning relevant prior distributions to all the parameters in the model, it is possible to calculate the posterior probabilities of all the regressors, as to whether or not they belong to the final model. The regressors with high posterior probabilities have a high likelihood of being determinants of variable under study, while those with low posterior probabilities are probably not significant as determinants. Overall, the results of Fernandez et. al. are in harmony with those of Sala-i-Martin. They find 18 regressors as being candidates for significant determinant, and their importance rankings of regressors have 95% Spearman rank correlation with those of Sala-i-Martin. However, there are some radical differences in the Bayesian approach to model selection, which involve a change in mindset and objectives of the entire procedure. These are discussed in the next paragraph.

The classical model selection problem is about identifying the determinants – finding the “true model”. In contrast, the Bayesian procedure provides posterior probabilities for each of the 2 trillion models. The best models are those which have posterior probabilities of only 0.1%; their chances of being true are around 1 in 1000. This is very high compared to 1 in 2 trillion, which is what models start out with. At the same time, it does not lead to much confidence in the model, especially when there are thousands of models with similar probabilities. Instead of attempting to identify a true model, the Bayesian method averages over all possible true models to provide inference and to provide forecasts. There are some important criticisms of Bayesian “Model Averaging” as opposed to Classical “Model Selection”. One of them is that we take average prediction over thousands of false models, in

which there is only one true model – how can an average over bad models lead to good results? The second is that the purpose of our research is to find out which variables matter and which don't, and in particular to find the right model. The Bayesian method gives us probabilities, which provide clues, but do not answer the question being posed. At the end of the analysis, we have thousands of models which have some claim to being the possibly true model. This does not help us much in achieving our goal, though it does provide some relevant information. It is also true that Bayesian techniques are sensitive to specification of the prior. The ones which are easiest to calculate with, the natural conjugate priors, turn out to give bad results in practice – see Zaman (1996, Section 3.4) for details. It is possible that a better prior, which may be more difficult to handle computationally, will yield better results.

Papers by Diebold (1989) and Hendry and Clements (2004) provide a way to reconciling these two conflicting approaches to choosing models. There are two different goals, and different strategies are suitable for achieving different goals. If we want to learn what the true model is, then the Encompassing methodology discussed in the next section is the far more effective strategy. However, if we want to make good forecasts, taking into account our uncertainty about the true model, than forecasts based on averages over diverse and conflicting models is the better strategy. Selecting the true model and then forecasting based on our selection corresponds to putting all our eggs in one basket. This is a high risk strategy. However, finding the true model is separately and directly of interest, in terms of finding out the mechanisms which govern the real world; here averaging over models is not helpful.

## **5. GENERAL TO SIMPLE MODELING**

As discussed earlier, in situations where there is a lot of uncertainty about determinants, the chances of hitting the exactly correct set of regressors are extremely small. In this case, it appears that an exploratory approach would be superior to a confirmatory approach; this leaves conventional econometrics behind, and is better done using various types of data visualization techniques which have made a lot of progress since computer software has become available to permit graphical examination of large data sets. In fact, I would strongly encourage students to learn these emerging techniques, which are likely to revolutionize the practice of data analysis. We turn to consider the other situation, where theory and empirical evidence suggests that there are a limited number of variables to examine. This is a situation where the strengths of Hendry's methodology stand out. This methodology was first introduced in Davidson et al. (1978); we will refer to this paper as DHSY. Gilbert (1986) provides a good discussion of the methodology and its advantages over conventional methodology. We will only discuss one aspect of this methodology, directly relevant to the topic under discussion. Instead of trying different combinations of variables, or searching over models, the General-to-Simple strategy asks us to start with the largest possible model, including all possible relevant regressors, and then simplify by excluding insignificant regressors. Obviously, including all possible regressors does take care of the possibility of omitted variables, assuming that data on all relevant variables is available. We now discuss the logic and merits of this procedure, which is a part of the general methodology of Encompassing developed and advocated by DHSY.

### **5.1. Why Simple-To-General Strategy Fails**

Conventional econometric methodology uses the simple-to-general approach. This means that we start with the simplest possible regression model which provides a reasonable fit to the data. If for some reason we find the model inadequate, then we attempt to find the simplest

possible extension of the model which will cope with the difficulties being encountered. For instance, a model might be inadequate because it forecasts poorly, and a solution might be found by adding a suitable variable. Even this seems like a simple, natural and logical procedure, it runs into a difficulty discussed in detail in the next paragraph.

A variable  $W$  can appear to be significant because it is acting as a proxy for omitted variables. Sala-i-Martin (1997) checked for this possibility by adding missing variables in groups of 3, and assessing whether  $W$  remains significant in 95% or more such additions. This is a simple to general strategy, since we start with a simple model, and assess the significance of  $W$  by adding 3 variables. We never consider the general model, which uses all the variables. Even though this strategy requires running two million regressions, it remains unsatisfactory for many reasons. One simple reason is that it is theoretically possible that considering groups of 4 variables may render  $W$  completely insignificant. Another reason is that simulations show that this strategy can fail to pick out the true models, making both type I and type II errors with high probability. That is, it can pick up irrelevant variables as significant, and discard relevant variables as being insignificant. Simulations showing this failure of Extreme Bounds Analysis, and variants of this developed by Sala-i-Martin have been carried out by Hoover and Perez (2004).

## **5.2. The Virtues of Starting with the GUM**

While omitted variables leads to seriously biased estimates, including irrelevant regressors does not lead to bias. This is the idea at the heart of the General To Simple (GeTS) methodology. We should start with the biggest possible model, which includes all potentially important regressors. This is called the General Unrestricted Model (GUM). Then we drop insignificant variables in order to arrive at a simple model. This is the reverse of the usual simple-to-general methodology, as we have explained earlier. We examine how it works in the context of the first three models for GNP we have already examined earlier in the paper.

One reason that the General-To-Simple approach has been avoided is that if we put in a very large number of regressors, most regressors appear insignificant. Furthermore, large amount of multicollinearity can be present among the regressors, and this makes it very difficult to pick out the right regressor from among competing possibilities. This was demonstrated by Lovell (1983), who showed that step-wise regression procedures frequently failed to converge to the right model, when started from a large model. However, using more sophisticated search techniques as well as newly developed methods for evaluating model selection procedures, Hoover and Perez (1999) come to the conclusion that model selection procedures for reducing the GUM to a simple model work reasonably well, approximately achieving standard significance levels for type I error, and respectably low rates of type II error. That is, if a variable is flagged as significant, then this is a mistake only in 5% of the cases. Furthermore, depending on the particular model being estimated, significant regressors are rarely excluded from the final model selected. This can fail to be true if the significant regressors exhibit very low variation, or strong multicollinearity, in which case the data does not provide strong evidence about their significance, and failures of statistical algorithms to pick relevant regressors can be expected.

Stripped of complications, the basic procedure for reducing the GUM is very simple. First consider the case where regressors are orthogonal (independent). Then each regressor gets a  $t$ -value and an associated significance level. Model simplification then involves simply dropping regressors with the lowest  $t$ -values from the regression. When regressors are

correlated, the situation becomes more complicated. When two regressors both have low  $t$ -values, it is no longer necessarily the case that the one with the lower  $t$ -value is the less important regressor. The search algorithms recognize this possibility and try multiple searches, starting from any one of the less significant variables. Each search terminates with a final simplified model. All of the final simplified models are considered together and tested against each other to select a single best final model. This entire process has been automated and is now available in a computer program called PC-GeTS (acronym for General to Simple modelling). Given a large list of variables, the program automatically reduces the regression model to the best simple model which omits all insignificant regressors, while retaining all significant ones. Of course there can be type I and type II errors in this process. Significant regressors may be omitted from the final simplified model, while insignificant ones may be retained. Nonetheless, these errors are controlled at reasonable levels, so that there is a chance of arriving at a good approximation to the true model. Thus The Axiom of Correct Specification requires that all relevant regressors must be included in a valid regression model. This is hard to manage using conventional simple-to-general methodology, since many relevant variables are not considered for inclusion. The general-to-simple approach starts by including all relevant variables and then simplifies by eliminating all insignificant variables. There are a number of complications that arise in the process. State-of-the-art algorithms for handling these problems are currently embodied in the PC-GeTS program, which provides automatic model selection procedures to select relevant and important regressors from a large list of candidates. Thus, it represents a solution to the problem under study in this note – how can we find a valid regression model, which does not omit any significant and relevant regressors. There is a very important caveat here; the General Unrestricted Model (GUM) must have all potentially relevant regressors. If there are important factors for which data is not available, then these cannot be included. Also, some important factors may not be known to the modeler. Subject to these cautions, the model selection algorithms provide a potential solution to the problem of nonsense regressions created by the Axiom of Correct Specification. Next, we examine how this works in practice, in the context of models for economic growth already examined earlier.

### **5.3. We Ran Just One Regression**

Hoover and Perez (2004) and also Henry and Krolzig (2004, “We Ran Just One Regression”) utilize the Automatic Model Selection strategies of PC-GeTS to re-analyze the growth regressions of Sala-i-Martin (1999). Because of missing data issues, they restrict analysis to a smaller set of 42 variables, which was analyzed earlier by Fernandez et. al. (2001), using Bayesian methods. Using a simulation study, they find that the methodology of Sala-i-Martin leads to the identification of too many regressors as significant; this parallels the performance of the procedure, which picks out 22 regressors as being significant determinants of growth. However the General-to-Simple procedure implemented in PC-GeTS has a good performance in terms of its ability to select the significant regressors. The procedure is not entirely automatic because we can choose the level of type I error; this is the probability of wrongly selecting a variable as being significant. One strategy, which allows for a 5% error and therefore chooses more variables, is labeled as the Liberal strategy. The other strategy is called Conservative; this allows for only 1% error, and chooses fewer variables. Utilizing a conservative strategy, Henry and Krolzig (2004) find that only three variables are significant in these growth regressions: (1) YrsOpen: the number of years that the economy was “open”; (2) EquipInv: Investment in Equipment; and (3) Confucius: The percentage of population following Confucian religion. A slight change in the level to match the settings of model selection procedures used by Hoover and Perez (2004) leads to the inclusion of (4) RevCoup:

Revolutions, Coups, and Wars, and (5) The percentage of Protestants in the country, as significant determinants of growth.

We have discussed several different procedures for selecting the relevant set of regressors from out of a large group of potential determinants. As we have seen, different procedures identify different variables. Is there a way to assess which of these different methodologies is more reliable? Hoover and Perez (2004), Hendry and Krolzig (2005), and Castle, Doornik, and Hendry (2011) address this problem by using simulations where there is a known true model. This allows us to assess different strategies for selection of regressors, by checking whether or not they are able to select the true regressors, known to us in the simulation. These simulations establish the superiority of the General-To-Simple approach, which has small type I and type II error probabilities, and tends to select the correct set of regressors. In contrast, EBA and Bayesian procedures have high type I and II error probabilities, and often select wrong variables, and leave out the true variables. However, as discussed earlier, the Bayesian procedures are not designed to select the true model; they are based on the concept of model averaging, which is an entirely different objective from model selection. When compared on the grounds of forecast performance, both Bayesian Model Averaging as advocated by Fernandez et. al. and the Automatic Model Selection procedures of GeTS, perform equally well, as reported in simulations by Bellone and Michaux (2006). Thus we arrive at the same conclusion reached earlier in Section 4.3: Bayesian model averaging and Model Selection based on encompassing are designed to solve different problems, and each procedure does well with respect to its intended objective. Comparing them on the same target is not justifiable and leads to misleading conclusions. Model averaging is good for forecasting, and insures against lack of knowledge of the model, while Model Selection puts all its eggs in one basket, and tries to guess select the right model out of trillions of possibilities. Each does fairly well at its own objective.

#### 5.4. Theory versus Empirics

A very important issue which has been neglected so far is that model selection is not a mechanical procedure – even though the automatic selection of models via PC GeTS makes it appear to be so. In fact, in order to evaluate model selection procedures, and to compare them in simulations, it is necessary to have a mechanical version. There are also many other contexts where automatic selection of regressors is desirable. Nonetheless, real world applications necessarily involve utilizing prior knowledge, historical and qualitative information, and other localized contextual data, to guide the process of selection of regressors. We explain and illustrate. Recall that regression of Pakistan Consumption on Honduras GNP leads to a nonsense regression, where the Honduras *GNP* appears as a highly significant regressor. As we explained earlier, this was due to the omitted variable of Pakistan *GNP*. If we put in this omitted variable, we get the following results:

$$\text{CONS(Pak)} = 0.225 + 0.69 \text{ GDP(Pak)} + 2.40 \text{ GDP(Hond)} + \varepsilon \quad (R^2 = 0.998)$$

(1.35) (0.06) (0.76) (2.38)

Normally, in these situations, a variable which acts as a proxy for an omitted variable becomes insignificant when the genuine variable is added. Here, even after the addition of Pakistan *GDP*, the Honduras *GDP* remains highly significant as a determinant of Pakistani consumption. A mechanical approach might lead us to conclude that *GDP* of Honduras is indeed a significant regressor. However, knowledge of real world suggests that it must be acting as a proxy for some other missing variable. So we would not be content to present above model as a final model; rather we would search for the missing variables which are

causing *GDP* of Honduras to be significant. This search would be guided by our knowledge of the real world, and not by statistical considerations.

Similarly, there are many cases reported in the literature where two regressors  $V$  and  $W$  are multicollinear, and  $V$  becomes insignificant in the presence of  $W$ . Thus statistics supports choice of  $W$  over  $V$ , but theory supports  $V$  over  $W$ . In such cases, we should prefer to be guided by theory, rather than by statistics. For instance, in the DHSY paper, the constant term is significant, while the theoretical error-correction term is not. DHSY go against statistical evidence and drop the constant, which make the error-correction term significant, and makes the equation more theoretically appealing. This leads to better forecast performance. This is because over the period under study, the variation in the error-correction term was low, which meant that the data was unable to recognize the significance of the term. In the long term, regressors which are theoretically significant will prove to be so empirically as well. Unfortunately, we always deal with finite amounts of data, where asymptotic properties do not always hold.

## 6. CONCLUSIONS

A long time ago, Leamer commented on the remarkable difference between the methods taught in textbooks, and techniques used in practice in econometrics:

I became interested in methodological issues as a University of Michigan graduate student from 1967 to 1970, watching the economics faculty build an econometric macro model in the basement of the building (The Michigan Model), and comparing how these same faculty members described what they were doing when they taught econometric theory on the top floor of the building. Though the faculty in the basement and on the top floor to outward appearances were the very same people, ascending or descending the stairs seemed to alter their inner intellectual selves completely.

The words “specification search” in my 1978 book *Specification Searches* refers to the search for a model to summarize the data in the basement where the dirty work is done, while the theory of pristine inference taught on the top floor presumes the existence of the model before the data are observed. This assumption of a known model may work in an experimental setting in which there are both experimental controls and randomized treatments, but for the non-experimental data that economists routinely study, much of the effort is an exploratory search for a model, not estimation with a known and given model. The very wide model search that usually occurs renders econometric theory suspect at best, and possibly irrelevant. Things like unbiased estimators, standard errors and  $t$ -statistics lose their meaning well before you get to your 100th trial model.

The differences between theory and practice noted by Leamer persist today. For an insightful discussion, see Kennedy (2002). The issues highlighted in the paper are well-known to professional econometricians, but rarely discussed in textbooks. The implications of misspecification for the practice of econometrics are routinely ignored in practice. As a result, students learn to do very sloppy econometrics, accompanied by very sloppy thought processes about econometrics. Two major steps are involved in going from data to understanding real



world economic relationships, and making policy recommendations. These are discussed below in separate sections.

### **6.1. Correctly Capturing Correlation Structures**

If two time series are stationary, then significant correlations between them signal the presence of some relationship. Unfortunately, the same is not true for observed time series of economic variable. Pick any two variables at random from the WDI data series, and you are like to observe a statistically significant correlation. Most of these correlations are spurious, while some are genuine. A genuine correlation is one which remains stable in a changing environment. Constructing a valid regression model requires substantially more effort, thought care and testing than is routinely done by aspiring econometricians. See Gilbert (1986) for a brief and clear description of the numerous steps involved. In this article, we have focused on only one necessity, that the set of regressors is correctly specified. Much more is involved in arriving at a valid regression model, which correctly captures stable correlational structures of a given data set. For more discussion of common methodological mistakes in econometrics, see Zaman (2012, “Methodological Mistakes and Econometric Consequences”).

### **6.2. Going from Correlation to Causation**

Far more difficult problems arise in converting an observation of patterns of correlation to causation, which is required for making policy statements. For instance, the percentage of Confucians in the population appears as a significant determinant of economic growth robustly, across a wide range of regression models. Does this mean that measures to propagate Confucianism would lead to higher growth? Almost surely not. Rather, this correlation between religion and growth only reflects the remarkable growth performance of China over the past few decades, and has no implications about causal links between the two. Converting correlation patterns to causal relationships is a delicate and difficult task, on which a lot of progress has been made in the last few decades. For some discussion of the relevant literature, see Zaman (2010, “Causal Relations via Econometrics”). One essential insight is that discovering causality cannot be reduced to a mechanical task. It almost always involves qualitative knowledge not easily captured in numbers. Finding relationships which persist through periods of structural change is a very important tool which aids the discovery of causal structures. For some guidance on how this can be done, see Asghar (2011) and Varian (2016). There have been many new and exciting developments in causal inference, which lies at the heart of requirements for policy built on regression models. Many of the important ones have been described in a new textbook by Angrist and Pischke (2014). But there will be some time lag before training in the classroom catches up, especially because the teachers trained in old-style econometrics are unfamiliar with these new methods.

**Final Words:** The goal of this pedagogical note is not to bring the applied econometrician up-to-date on methodology. There have been a huge number of important and relevant developments in the past few decades, which would take many books to cover. Rather, our goal has been to focus on one important and common methodological mistake made by many applied econometricians. The discussion has been framed around the General-to-Simple methodology, which is one of the many useful and important methodological prescriptions developed in the DHSY paper in the 1970’s. This is sometimes called the Hendry methodology or the LSE methodology, and has been debated and discussed extensively. For a

recent critical evaluation of the strengths and weaknesses of this methodology, see Faust and Whiteman (1997).

## REFERENCES

- Angrist, J.D. and J.-S. Pischke (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press.
- Asghar, Z. (2011). A Structural Approach for Testing Causality. *International Econometric Review (IER)*, 3 (2), 1-12.
- Bellone, B. and E. Michaux (2006). Model Uncertainty and Forecasting, a Practitioner Point of View. *Unpublished manuscript*.
- Castle, J. L., J.A. Doornik and D.F. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics*, 3(1).
- Davidson, J. E., D.F. Hendry, F. Srba and S. Yeo (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *The Economic Journal*, 661-692.
- Diebold, F.X. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5 (4), 589-592.
- Faust, J. and C.H. Whiteman (1997). General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: A translation and critique. *Carnegie-Rochester Conference Series on Public Policy*. 47.
- Fernandez, C., E. Ley and M.F.J. Steel (2001). Model uncertainty in cross-country growth regressions. *Journal of applied Econometrics*, 16 (5), 563-576.
- Freedman, D. A. (2010). *Statistical models and causal inference: a dialogue with the social sciences*. Cambridge University Press.
- Gilbert, C.L. (1986). Practitioners' Corner: Professor Hendry's Econometric Methodology. *Oxford Bulletin of Economics and Statistics*, 48 (3), 283-307.
- Granger, C.W.J. and H.F. Uhlig (1990). Reasonable extreme-bounds analysis. *Journal of Econometrics*, 44 (1-2) 159-170.
- Hendry, D.F. (1995) *Dynamic Econometrics*. Oxford University Press, Oxford.
- Hendry, D.F. and M.P. Clements (2004). Pooling of forecasts. *The Econometrics Journal*, 7 (1), 1-31.
- Hendry, D.F. and H.-M. Krolzig (2004). We ran one regression. *Oxford bulletin of Economics and Statistics*, 66 (5), 799-810.

- Hendry, D.F. and H.-M. Krolzig (2005). The properties of automatic Gets modelling. *The Economic Journal*, 115 (502).
- Hlavac, M. (2016). ExtremeBounds: Extreme Bounds Analysis in R. *Journal of Statistical Software*, 72.
- Huff, D. (1954) *How to Lie With Statistics*, Norton, NY.
- Hoover, K.D. and S.J. Perez (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *The Econometrics Journal*, 2 (2) 167-191.
- Hoover, K.D. and S.J. Perez (2004). Truth and robustness in cross-country growth regressions. *Oxford bulletin of Economics and Statistics*, 66 (5), 765-798.
- Kennedy, P.E. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys*, 16 (4), 569-589.
- Leamer, E.E. (1978). *Specification searches: Ad hoc inference with nonexperimental data*. Vol. 53. New York: John Wiley & Sons Inc.
- Lovell, M.C. (1983). Data Mining. *Review of Economics and Statistics*, 45, 1-12.
- Sala-i-Martin, X. (1997). I just ran two million regressions. *The American Economic Review*, 178-183.
- Stock, J.H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica: Journal of the Econometric Society*, 1035-1056.
- Tukey, J.W. (1977). *Exploratory data analysis*. New York: Wiley.
- Varian, H.R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences*, 113 (27), 7310-7315.
- Zaman, A. (1996). *Statistical foundations for econometric techniques*. Academic Press.
- Zaman, A. (2010). Causal Relations via Econometrics. *International Econometrics Review*, 2 (1), 36-56.
- Zaman, A. (2012). Methodological Mistakes and Econometric Consequences. *International Econometric Review*, 4 (2), 99-122.