*Iranian Economic Review, Vol.6, No.6, 2002*

# Scientific vs. Cookbook Econometrics An emphasis on the Ethical Issues

By:
Ahmad Jafari-Samimi, Ph.D.*

ABSTRACT

During the 1960's, many as was firmly supported by the historical founders of econometrics, had hoped that econometrics would provide a sound scientific foundation for econometrics in which each element of specification would be determined primarily on the basis of economic theory. However, due to misusing of econometrics and also wide usage of the so called cookbook econometrics, many researchers - especially economists - have recently lost some confidence in econometrics.

The present paper discuss the ingredients of the scientific econometrics and suggest ways of helping to renew optimism about scientific econometrics or as I have called it "scienceometrics" with an emphasis on the ethical issues.

**Key words:** scientific econometrics, cookbook econometrics, scienceometrics, and ethical issues.

---

**- This paper was written during the author's sabbatical visit (July 2000-January 2001) at The University of Newcastle, Australia.*

*Professor of Economics, University of Mazandaran at Babolsar, Iran.*
*E-mail: jafarisa@yahoo.com*

## 1. Introduction

Many economists- including myself-appear to have recently lost some confidence in econometrics. The main reason is that econometrics has been misused by many researchers, especially in non-economics fields. They have manipulated the data (using a specific sample of sub-sample), the variables (using different versions), lag-structures (using different time-lags), mathematical forms of the model (linear or non-linear), estimation procedures (weighted instead of ordinary least squares), and so on without any scientific justification and then to choose the equation that conforms the most to what they want the results to look like. This way of dealing with econometrics and model selection which may truly be called "cookbook" econometrics is completely opposed to the scientific econometrics in which each element of specification should be determined primarily on the basis of economic theory rather than on the results of a desired estimated regression equation.

During the 1960's, many had hoped that econometrics would provide a sound scientific foundation for economics. During this period, economists sought to influence government policy not only through economic theory but also through the wide availability of data and quantitative information.

By the 1980's and 1990's and even today, we are somewhat witnessing a renewed optimism about the role of econometrics and it could be argued that econometrics is not merely considered as a toolkit for economists, but a subject which has been used by merely all researchers in other fields such as accounting, management, sociology, psychology, law, medicine, and so on in such a way that I would suggest the term of "scienceometrics" as a suitable substitute for econometrics. However, I think we should agree that today econometrics as a branch of science in the introduction of scientific tools and methods into economics is suffering from the ethical point of view.

The main purpose of the present paper is to discuss the ingredients of the scientific econometrics and suggest ways of helping to renew optimism about scientific econometrics with an emphasis on the ethical issues. The paper proceeds as follows: in section II, we briefly review the history of econometrics as a background to see how much emphasis the founders of econometrics put on the scientific methodology of econometrics. Section III deals with the main ingredients of the scientific as compared to the so-called cookbook econometrics with some important examples. Finally section IV summarizes the concluding remarks with an emphasis on the ethical issues in econometrics.

## 2. A Brief History of Econometrics

Empirical analysis in economics has a relatively long history; its origins can be traced as far back as the sixteenth century when the "political arithmeticians" led by Sir William Petty analyzed problems such as taxation and international trade with quantitative information[1] However, econometrics, as we currently understand the term, is of much more recent origin, and is marked by the foundation of the "Cowels Commission"[2] and the Econometric Society in the 1930's[3]. One important characteristic of macromodel building at the Cowles Commission was the association of statistics with economics. The specification and application of economic models had to be firmly grounded in received or newly developed economic analysis. This method impressed Laurence Klein in his first professional position, right out of graduate school, that this was the proper way to proceed[4].

Early econometric studies in Europe came from several institutions. R. Frisch, one of the founding fathers of econometrics, built up his econometric project at the Oslo Institute of Economic Research. He guided a number of now well-known econometricians at the Institute, such as T. Haavelmo. The Dutch Central Bureau of Statistics and the Netherlands Economic Institute were Centres of Macroeconometric Models. The research was pioneered there by J. Tinbergen, another founding father of econometrics.

It should be noted that prior to the 1930's, the Probability theory was commonly rejected as being unsuitable as a basis for analyzing economic data. Change took place from the 1930's as more and more statistical methods were tried in the field of applied economics. The changes were also accelerated by the substantial progress made in mathematical statistics with regard to multivariate models. However, at that time the contributions of econometrics to economics were little and slow[5]. By the time that T. Haavelmo argued for a full adoption of the Probability approach as the foundation of econometric theory in the early 1940's, -later referred to as the "Haavelmo Revolution", things did not change much in econometrics[6] The concept of a "Model" did not exist at the time of earliest works. During the pre-model period, econometric practice was merely concentrated on measuring approximately certain economic laws by descriptive statistics. When the model had been created, it became as the vehicle to comfort, with data, economic theory specified in a measurable and testable equation.

Models thus made up the essence of econometrics.

It should be noted that the statistical method was assumed to have been looked after by statisticians who had introduced estimation methods such as the ordinary least squares (OLS). Simple regressions by OLS method was used widely. The limitations of OLS and problems faced by econometrics such as identification, mulitcollinearily, hetroscedasticity, measurement error, causality, non-stationarity, cointegration and so forth were discussed by econometricians during the last three decades or so.

## 3. Ingredients of the Scientific Econometrics

Scientific econometrics is based on the methodology of econometrics. According to Machlup[8] methodology is the study of the principles of discrimination that guide researchers in deciding whether to accept or reject propositions as a port of the body of ordered knowledge in their own discipline. In addition, as DeMarchi and Gilbert[9] put it "...methodology is inquiring into why the accepted is accepted".

When the term "econometrics" was first used it conveyed both the development of pure economic theory from a mathematical viewpoint, and the development of empirical estimation techniques for economic relationships. This is virtually reflected in the constitution of the Econometric Society, which defined the society's primary objective as "the advancement of economic theory in its relation to statistics and mathematics[10].

All scientific definitions of econometrics are based on economic theory or economic analysis and the substance of economic is more important than the statistics or mathematics. For example Wassily Leontief's[11] definition of econometrics as "a special type of economic analysis in which the general theoretical approach often formulated in explicitly mathematical terms", or the definition given by Samuelson, Koopmans and Stone[12] "as the quantitative analysis of actual economic phenemona" are supporting our point.

Scientific econometrics is an important tool which economists and now other researchers may use for their own purposes. It can test different hypotheses concerning economic or non-economic theories. Without economics or science, econometrics or scienceometrics simply becomes metrics, though even then it would rely on some theoretical language and justification to define and estimate the models of interest[13].

The most important stage in applied scientific econometrics is therefore, the specification of the theoretical model and its scientific theoretical justifications rather than selecting an ad-hoc model without theory. After choosing the dependent variable, the following components should also be specified as the ingredients of the scientific econometrics: the dependent variables based on theoretical reasoning and how they should be measured; the mathematical form of the equation and the type of error term in the equation. I should again insist that each of the elements of specification be determined primarily on the basis of economic or other science theories rather than on the results of an estimated regression equation.

It should be noted that from the scientific econometrics point of view the best models are those on which much cares has been spent to develop the theoretical underpinnings and as mentioned by Studenmund[14] only a short time is spent pursuing alternative estimations of that equation. We can therefore call scientific econometrics as good or ethical econometrics as compared to the so-called cookbook econometrics.

The following are only some examples of the widely seen misinterpretation and sometimes misusing of econometrics, which caused many researchers to lose confidence in econometrics:

1) Trying to select an estimation equation with a good fit (high $R^2$).

It should be noted that the goal of scientific econometrics and regression is not to maximize $R^2$, but unfortunately many researchers- log GNPP especially students and sometimes even graduate students, mostly in developing countries, find it hard to resist that temptation. We have to remember that the quality of fit of an estimated equation is only one measure of the overall quality of that regression. An estimated equation with a very low fit but with plausible and significant signs for estimated coefficients may be considered as a useful equation and should not be discarded because of low $R^2$. For example, the regression of a tax-ration on log GNP per capita by Burgess and Stern[15] in 1993 for the whole sample of 82 countries results:

$$\frac{T}{GDP} = 5.78 + 1.84 \log GNPP$$

$$(1.02) \quad (2.02)$$

$$\acute{R}^2 = 0.04$$

Where, the coefficients in parentheses are t- values. The finding is weak because of a very low value of $\acute{R}^2$ but significant (at the 5 percent level). The very low $\acute{R}^2$ indicates a wider scatter of points around the linear relationship between T/GDP and log GNPP[16].

2) Problem of model selection based on $R^2$.

It should be noted that it is not always a suitable decision to prefer a nonlinear to a linear equation or vice versa, merely base on the value of $R^2$. To be more clear, consider the following two equations:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + U_t \quad (1)$$

$$\ln y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \quad (2)$$

As it is seen the dependent variable in these models are not the same, therefore, their total sums of squares and as a result of that $R^2$ will be different[17].

3) Problem of causality.

Another important warning is that while it is true that many economic relationships are causal by their nature, but it should be reminded that the estimated regression equations, no matter how statistically significant can not show causality. In scientific econometrics it is mainly the economic theory and sometimes the researcher's own knowledge and common sense which suggest the causality. All regression equations show is to see whether a significant quantitative relationship exists. Even causality tests such as those introduced by Granger and Sims and others which are based merely on statistical and mathematical reasoning rather than economic justification, can only be used to show the direction of a causal relation and in scientific econometrics I suggest these tests should be used as a last resort regarding the direction of the relationship when there are ambiguities[18].

## 4. Concluding Remarks

The main important purpose of the present paper was to show that due to the wide usage of the so-called cookbook econometrics rather than the scientific econometrics-in which each element of specification would be determined primarily on the basis of economic theory and was firmly supported by the historical founders of econometrics-many researchers have lost confidence in econometrics.

Based on discussion given in this paper the following are some important suggestions to improve the renewed optimism about applying scientific econometrics:

1) Economists should use structural models that are characterized by the use of economic theories in specifying relations rather than the vector autoregressive models, which are really measuring without theory.

2) Never discourage researchers, especially students from having low $R^2$ in their estimated regression models. Never let them, as mentioned by Mayer, torture the data long enough until they confess[19].

3) Try to select the model, variables and the lags and so forth based on theoretical underpinning rather than on the estimated results. From the ethical econometrics point of view, try to run a few different specifications while attempting to avoid the major econometric problems and as mentioned by Studenmund, report honestly the number and type of specifications.

4) Never expect causality from the estimated regression equations. Use only causality tests as a last resort to solve ambiguities concerning certain economic relationships.

5) Due to the limitations regarding the estimated equations caused by problems such as inaccurate data, incorrectly formulated equations, improper estimating techniques, and so on, always show the regression results with caution[20].

6) Finally, it is suggested that the economists, rather than statisticians or others, teach econometrics. I think economists care more about the theoretical underpinning as well as the ethical issues of the proposed regression models than others[21]. I also suggest a section of the econometrics syllabus be devoted to the ethical issues of econometrics.

## Notes

1) Among the early forerunners of econometrics, in addition to William Petty, others such as Gregory King, Edegeworth, and Pareto are also mentioned. Edegeworth and Pareto, more than any other writers, by linking the three subjects of economics, statistics, and mathematics, virtually played an important role in introducing what is now called econometrics. See Fisher (1941) for details.

2) The Cowels Commission for research in economics, a non-profit corporation with close associations to The Econometric Society enjoyed a growing reputation during the 1930's as a center for mathematical

economics. It had been founded in Colorado Springs, USA in 1932 by Alfred Cowles III, an investment advisor and a member of a wealthy family, in the hope that the application of mathematical methods to the study of economic issues would lead to better predictions of stock market behavior. Cowles attempted to recruit economists such as Irving Fisher, Harold Hotelling, and Ragnar Frisch. Later leading economists and statisticians, such as RA Fisher, Joseph Schumpeter, Jacob Marschok, TO Koopmans, and L Klein joined the commission. For details on history of the "Cowles Commission" see Epstein (1987).

3) In addition, in the 1940's, the department of applied economics in Cambridge, UK was founded. See Darnell and Evans (1990) for the details. It should also be noted that there had been some attempt to form an econometric society by Fisher as early as 1912 but there were too few economists interested in such a society. See Christ (1985), Epstein (1987), Darnell (1984), and Morgan (1989) for more details on this issue and on the history of econometrics.

4) See Klein (1985), P. 11

5) It could be argued that the real acceleration in econometric studies is largely a post Second World War phenomenon. The availability of high speed computers and statistical packages and also the increasing access to economic data played an important role in high activity of econometric works in this period and onward.

6) See Duo (1993), Haavelmo (1944). A key message of Haavelmo's advocacy that the probability approach was the right vehicle not only for closing the gap between economic theory and econometrics, but also for changing economists' view point into a stochastic world.

7) Models were classified into theoretical and statistical ones soon after introducing the concept. The former was applied to models derived directly from a prior theory, and the latter to models formed through data-fitting. See Duo (1993) I bid p.37. Also more recently, econometric models were further divided into two new classes: structural models and vector autoregressive (VAR) models. The first class of models is characterized by the use of economic theory in specifying relations and accounting identities. Naturally, the specification takes the form of identifying restrictions. Vector autoregressive models do not impose parameter restrictions. In other words, they as mentioned by J Marquez, constitute a sophisticated version of "measuring without theory". Since these models do not have any theoretical ingredients, they can be used

only in forecasting. See Marquez (1985).

8) See Machlup (1978, PP. 54-55). It should be noted that these are not universally accepted definitions of methodology. See for example Blaug (1980) for details. However, in scientific econometrics, I think we should agree that the methodology of econometrics is a subpart of the methodology of economics. This is because econometrics is part of the economics, rather than a field of its own.

9) See De Marchi and Gilbert (1989, P. 9).

10) It should be said that the origin of the word econometrics is relatively new. Econometrics is a connection of two Greek words, "oikonomia" (administration, or economics) and "metron" (measure). However, it should be added that the origin of a word does not itself give us its meaning. For various definitions of econometrics, see Poirier (1994) volume 1,PP.9-13.

11) See G. Tintner "The Methodology of Econometrics" volume 1, edited by D.J. Poirier (1994), PP7-13.

12) See Samuelson, Koopmans, and Stone (1954) P.14. It should be noted that there is a group of definitions of econometrics in which stresses the statistical rather then the economical aspects. See for example Haavelmo (1944), and Lange (1945).

13) It should be noted that this view contradicts that of De Marchi and Gilbert that "econometrics is now a fully fledged and distinct discipline, lying between mathematical statistics and economics, drawing on the one and indispensable to the other". See De March and Gilbert (1989) P.11.

14) See Studenmund (1992).

15) See Burgess and Stern (1993). It should be noted that cross-section studies usually have low $\acute{R}^2$ compared to time series studies where most variables show some trends and there are high correlation among them.

16) I would like to remind readers, especially students, that the above model with a very low $\acute{R}^2$ has been published in one of the well-known scientific journal. Therefore, it is not always right to put too much emphasis on choosing a model with high $R^2$ Sometimes a higher $R^2$ model may result an unexpected sign in some of the variables. Thus a researcher who uses $R^2$ as the only measure of the quality of an equation (at the expense of economic theory) increases the chances of having misleading results. See for example Studenmund (1992) Ibid,

P.53.

17) We can write:

$$TSS_1 = \Sigma\,(Y_t - \bar{y})^2$$

$$TSS_2 = \Sigma\,(\ln Y_t - \frac{\Sigma \ln Y_t}{m})^2$$

So it is improper to compare the value of $R^2$ from these two models because their dependent variables are different. $R^2$ in a linear model measures the percentage of variations in Y explained by the independent variables, whereas in a log-linear model, $R^2$ measures the fraction of variation in lnY explained by the model. One way to get around this problem is to compute a new comparable $R^2$ or as it is called "quasi-$R^2$" for the nonlinear model by transforming the predicted values of the nonlinear dependent variable into a form that is directly comparable to the original dependent variable. In other words, we can write:

$$\text{Quasi } R^2 = 1 - \frac{\Sigma\,(Y_t - \text{anti-log}(\ln \hat{Y}_t))^2}{\Sigma\,(Y_t - \bar{y})^2}$$

See Studenmund (1992) I bid, P. 228 and Ramanathan (1995) for the details. It should also be noted that in the last two decades or so, several criteria for choosing among models have been introduced. All of these take the form of the residual sum of squares multiplied by a penalty factor that depends on the complexity of the model. For example Akaike (1970, 1974) developed two methods, one known as the finite predictor error, and the other known as the Akaike information criterion. Other criteria includes those by Schwartz (1978), Hannan and Quinn (1979), Shaibata (1981), Rice (1984) and a generalized cross validation method developed by Craven and Wahba (1979) and used by Engle, Granger, Rice, and Weiss (1986). Statistically, one may prefer a model that outperforms another in some of these criteria.

18) Ambiguities with respect to a one-way or two-way relationship and also with regard to distinction between dependent and independent variable. For example, causality tests could be applied to determine the direction of the relationships between inflation and budget deficit or investment and growth, which are theoretically unknown. It should also be noted that causality tests such as Granger or Sims, only takes as its Premiss

that the past can cause the future, the future can not cause the past and such a test of causality does not allow any particular philosophical position to be adopted regarding the causal structure of the Y, X relationship. So it is important to note that in these tests what is actually being tested is more a temporal ordering and a predictive ability rather than "causality" as the word is commonly understood in the science.

19) See Mayer (1980), P.175.

20) It should be noted that despite some progress made recently in testing problems such as non-stationary and cointegration, I should remind readers that even these new tests suffer from some limitations. See for example Hakiko and Rush (1991) for details on the difficulties of detecting cointegration over short time periods.

21) I agree with Marquez who said, "an econometrician wears two hats. In formulating behavioral relations, we wear a theorist's hat since we assume the parameters of behavioral relations to be known. In estimating the parameters, we wear a statistician's hat since we take the behavioral relations as given". See Marquez (1985), P.1.

## References


1. *Akaike, H. (1970). "Statistical Predictor Identification". Annals Institute of Statistics and Mathematics, Vol.22,PP,203-217.*
2. *Akaike, H. (1970) "Statistical Predictor Identification". IEEE Trans, Auto. Control, Vol.19, PP.716-723.*
3. *Blaug, M. (1980). The Methodology of Economics, or How Economists Explain, Cambridge, Cambridge University Press.*
4. *Burgess, R., and N. Ster (1993). "Taxation and Development", Journal of Economic Literature, Vol.31. June 1993, PP.762-830.*
5. *Christ, C.F. (1985). "Early Progress in Estimating Quantitative Economic Relationships in America", American Economic Review, Vol.16, PP.54-74.*
6. *Craven, P. and G. Wahba (1979). "Smoothing Noisy Data with Spline Functions" Numerical Mathematics, Vol.13, PP.377-403.*
7. *Darnell, A.C. (1984). "Economic Statistics and Econometrics", in O'Brien, D.P and J. Creedy (eds) Economic Analysis in Historical Perspective, London, Butterworth, PP.152-182.*
8. *Darnell, A.C., and J.L Evans, (1990). "The Limits of Econometrics", Edward Elgar.*
9. *De Marchi, N., and C. Gilbert, (eds). (1989) History and Methodology of Econometrics, Oxford Economic Papers, 41.*

10. *Duo, R. (1993). The Formation of Econometrics-A Historical Perspective, C Larendon Press, Oxford.*

11. *Engle, R.F., C.W Granger., J. Rice., and A. Weiss (1986) "Semi-Parametric Estimates of the Relation Between Weather and Electricity Sales", Journal of American Statistical Association, Vol.81.*

12. *Epstein, R.J. (1987). A History of Econometrics, North Holland.*

13. *Fisher, I. (1941). "Mathematical Method in Social Sciences", Econometrica, Vol.9, July-October, PP.182-197.*

14. *Haavelmo, T. (1944). "The Probability Approach in Econometrics", Econometrica, Vol.12, Supplement.*

15. *Hakiko, G.S., and M. Rush (1991). "Cointegration:How Short is the Long Run", Journal of International Money and Finance, PP.231-254.*

16. *Hannan, E.J., and B. Quinn ( 1979), "The Determination of the Order of an Autoregression", Journal of Royal Statistics Society, Series B, Vol.41, PP.190-195.*

17. *Klein, L. (1985). Economic Theory and Econometrics, edited by J. Marquez, Basil Blackwell.*

18. *Lange, O. (1945). "The Scope an Method of Economics" Review of Economic Studies, Vol. 13, PP.19-32.*

19. *Machlup, F (1978). Methodology of Economics and other Social Sciences, New York: Academic Press.*

20. *Marquez, J. "Forty Years of 'Rigorous Observational Positivism'" in Economic Theory and Econometrics, L. Klein, Basic Blackwell 1985, PP.1-10.*

21. *Mayer, T. (1980) "Economics as a Hard Science: Realistic Goal or Wishful Thinking?" Economic Inquiry, P. 175.*

22. *Morgan, M.S. (1989).The History of Econometric Ideas, Cambridge, Cambridge University Press.*

23. *Poirier, D.J. (1994). "The Methodology of Econometrics", Vol.1, Edward Elgar.*

24. *Ramanathan, R. (1995). Introductory Econometrics with Applications, 3rd Ed, Fort Worth: Dryden Press.*

25. *Rile, J. (1984). "Bandwidth Choice for Non Parametric Kernel Regression", Annals of Statistics, Vol.12, PP.215-230.*

26. *Samuelson, P., T.C. Koopmans., and J.R. Stone. "Report of the Evaluation Committee for Econometrics", Econometrica, 1954, P.141.*

27. *Shwartz, G. (1978). "Estimating the Dimension of a Model", Annals of Statistics, Vol.16.*

28. *Shibata, R. (1981). "An Optimal Selection of Regression Variables", Biometrika, Vol. 68.*

29. *Studenmund, A.H., and H.J. Cassidy (1992). Using Econometrics, A Practical Guide, Second Edition, Harper Collins Publishers.*