

AUTOMATED INFERENCE AND THE FUTURE OF ECONOMETRICS: A COMMENT

PAOLO PARUOLO
University of Insubria

This note discusses the (dis-)similarities between automated inference and computer-aided decisions, at the interface of econometrics and economics. It is argued that computer-aided decisions are best suited for scientific communication. For the future, the topic of learning is singled out as one of the most promising areas of integration of econometric techniques and economics.

It is a pleasure to participate in the discussion on Automated Inference and the Future of Econometrics at the 20th anniversary mark of *Econometric Theory*. Many factors appear to be fueling the growth of computer-intensive inferential rules; some of these lie with theoretical advances and with the current and predictable increases in computing power and in availability of large data sets.

Among the many possible aspects, I wish to comment on three issues related to this theme, with a view toward possible developments. These comments are not intended to highlight some published work of mine, and they are mostly non-technical; I hope none of these features will be seen as a liability by the reader.

The three issues, listed in increasing degree of importance, concern the distinction between automation and computer aided decisions (Section 1), model selection (Section 2), and learning (Section 3). Conclusions are reported in Section 4.

1. AUTOMATION AND COMPUTER AIDED DECISIONS

The first aspect is wording. I perceive automated inference (AI) as associated with artificial intelligence and expert systems. Many computing-intensive procedures are instead of a different nature, which I would classify as tools for computer aided decisions (CAD), to be defined subsequently. This section argues that the two concepts are different and that they may be fruitfully applied in different situations. In particular AI is best suited for industrial applications, CAD for scientific communication.

The phrase “automated inference” conveys the idea that control over inference is left to a computer program or, more precisely, that the econometrician

I thank Peter Phillips and an anonymous referee for useful remarks on a previous version of this comment. The usual disclaimer applies. Address correspondence to Paolo Paruolo, Dipartimento di Economia, Università dell’Insubria, via Ravasi 2, 21100 Varese, Italy; e-mail: paolo.paruolo@uninsubria.it.

does not have complete control over inference. This concept is associated with artificial intelligence and expert systems because whatever inferential rule is being applied, it is used without the direct control of the econometrician. In the following I will use the (common) acronym AI for both phrases.

Examples of AI are “black-box” procedures. Many black boxes have been entertained in econometrics, especially in the past. They include, e.g., automatic univariate autoregressive moving average (ARMA) modeling, i.e., completely human-unaided computer software that takes in data and that produces predictions using an ARMA model, selected via a set of inference rules. More recent references include some applications of Bayesian vector autoregressions with Minnesota priors, which are close to being black boxes.

Automated procedures of this kind are also in use for fitting and selecting artificial neural networks. Some selection methods, called “pruning,” are typically computed before the optimization (estimation) has been completed. These pruning procedures cannot be restated as (quasi-) likelihood ratios and thus are not amenable to direct statistical interpretation, at least using standard theory.

By contrast let us define by computer aided decisions (CAD) any computer intensive inference procedure that is (in principle) completely supervised by a human. Examples of CAD are abundant, including plug-in bandwidth kernel density estimation, spectral estimation and testing, indirect inference, Monte Carlo estimation, model selection through information criteria or statistical tests, and many more.

The common denominator to any definition of scientific methodology is communication. For experimental sciences, scientific validity is associated with replicability of experiments. For nonexperimental sciences, replicability of experiments is replaced by replicability of inference. To communicate, econometricians need to have complete control over inference, and CAD is the elective method for large data sets in macro- and micro-econometrics. A different situation exists in financial econometrics; see the discussion that follows.

In fact, there are few reasons why the econometrician should give away control over inference if the goal of the research project is scientific communication, despite a possibly heavy computational burden. If the key steps of inference are unknown to the econometrician, there is little hope she or he can convince the scientific community that the prompted inference is a sound one.

This reasoning connects with the aspiration of econometrics to help economics become more scientific; this idea is at the basis of the Constitution of the Econometric Society. Concerning the relation between science and inference, Kim, De Marchi, and Morgan (1995) define the following four motivations for testing in econometrics: (1) theory falsification, (2a) theory consensus building, (2b) model quality control, (2c) matching specific model characteristics with a subset of empirical data (my own wording). Number (1) corresponds to Popperian theory falsification, number (2) to theory confirmation.

Regardless of which school in philosophy of science one subscribes to, it is hard to see which of the preceding activities one can recommend to be left undisclosed, as in AI. For instance, in (2a), tests are performed to assess the

“accuracy and reliability of [theory] performance in order to secure widespread support for belief in that theory” (Kim et al., 1995, p. 83); obviously this message cannot be conveyed if inference (performed automatically) is not fully known.

On the other hand, one can see many advantages of AI in industrial applications, where the main goal is on timely decision making. The closest analogues to industrial quality control problems are found in finance. In applied finance, timely forecasts on a vast number of phenomena are needed to make buy/sell decisions, and automation is a must more than an option. For obvious reasons, financial analysts are also averse to communicating their inferences to competitors, and AI rules prevail. Given that financial decision makers probably use AI, this information could be used in model building; see Section 3.

2. MODEL SELECTION

Model selection has a long history in econometrics, possibly as long as the history of econometrics itself. Many procedures are currently in use for model selection, including information criteria (IC) and hypothesis testing (HT). Sometimes model selection is an intermediate step in an inference procedure, e.g., in a forecasting exercise. Other times model selection is the main goal of inference; this happens, e.g., when one wishes to compare several economic theories as alternative explanations for the same set of data. In this case, multiple comparison (MC) methods can be applied; these methods are not presently as widespread in econometrics as in other sciences. Sections 2.1–2.3 report some comments on IC, HT, and MC.

2.1. Information Criteria

There has been substantial progress in the area of IC over the last decades; see Rao and Wu (2001) for a recent survey. Progress has been made in extending results to possibly nonstationary time series processes; see Phillips (1996) and references therein for the extension of Rissanen’s theorem to the case of time-series prediction. IC, although based on (quasi-)likelihood ratios (see Pötscher, 1991), treat models symmetrically, unlike hypothesis testing. This is the key motivation used by Granger, King, and White (1995) in advocating the use of IC in comparison of economic theories.

One limitation of IC lies however with their characteristic of estimators of the best model. No measure of strength of decision is usually available, unless a Bayesian paradigm is used. Bayesian analysis is however not the best suited method for communicating inferences when the elicitation of a priori distributions is subjective. Model averaging is not an option when inference is aimed at selecting among economic theories.

In general, one would like to complement the choice of a model with a likely (objective) assessment of whether this model is significantly distant from alter-

native ones. If one is to choose among different economic theories, this is particularly relevant.

Asymptotic probability statements on IC can be deduced from their connection to HT and MC procedures, which are discussed in Sections 2.2 and 2.3. This is because IC, HT, and MC are all based on the same (quasi-)likelihood theory. No investigation appears to exist in this direction in the literature, and any attempt along this line will probably be rewarded.

One other area where more theoretical contributions would be welcome is comparison of IC procedures. These comparisons should be aimed at selecting a single or at most a few best criteria. The many IC now in use often generate conflicting decisions, with no clue on how to resolve them.

At present, procedures are classified as efficient and/or consistent (see Rao and Wu, 2001). Among consistent criteria, no comparison is usually made on (higher order) asymptotics, as in the case of parameter estimation.

Here one needs to define (higher order) asymptotics in this context. As one possibility, consider two criterion functions $h_{jn}(i), j = 1, 2$, based on a sample of size n , where $i = 1, \dots, p$ indicates different models M_1, \dots, M_p . Let i_0 be the index of the correct model. Consistency of $\hat{i}_{jn} := \arg \min_{1 \leq i \leq p} h_{jn}(i)$ is usually proved by showing that $h_{jn}(i_0) = O_p(1)$ as $n \rightarrow \infty$ whereas $h_{jn}(i) = O(n^{\alpha_j}) \rightarrow \infty$ for $i \neq i_0$. The order n^{α_j} could depend also on i , but we assume for simplicity that it is uniform over incorrect models. Assume that both h_{1n} and h_{2n} are consistent. One could first compare α_1 and α_2 , where the criterion that guarantees the fastest divergence should be favored.

Assume however that the two criteria h_{1n} and h_{2n} are both consistent and have the same divergence rate $\alpha_1 = \alpha_2$. One could then study the ratio $q_n := h_{1n}/h_{2n} - 1$ to ascertain if q_n is asymptotically positive or negative. More generally one could study the asymptotic distribution of q_n and prefer h_{1n} to h_{2n} if $\Pr(q_n \geq 0) \rightarrow \eta > \frac{1}{2}$. These comparisons could give a handle on how to select a best criterion, thus reducing the possibility of conflicting results.

2.2. Hypothesis Testing

Some selection procedures are based on HT. HT gives a probabilistic framework within which one can measure strength of decisions. Let $\Theta_1 \subset \Theta_2 \subset \dots \subset \Theta_p$ be nested parameter sets corresponding to models M_1, \dots, M_p . Some common misconceptions on HT in model selection are the following: (i) in nested models, a general to specific HT strategy is only associated with testing M_{j-1} in M_j starting with $j = p$ and proceeding to $j = 0$ (called “testing-down” strategy in the following discussion); (ii) HT may be applied only to nested models that contain the true data generating process (DGP).

Consider (i). In several instances it has proved useful to consider a general to specific strategy that tests M_i in M_p starting with $i = 0$ and proceeding up to $p - 1$. For later reference, call this procedure a “testing-up” sequence. This strategy is still general to specific, because each submodel is tested against the full model M_p .

The testing-up sequence has been applied, e.g., when testing for cointegration rank (see Johansen, 1992). Here M_j indicates a vector autoregressive (VAR) model with cointegration rank at most equal to j ; let $LR(i|j)$ be the likelihood ratio test comparing M_i and M_j . For the testing-up procedure, as sample size diverges, one has $\Pr(\text{select } M_r) \rightarrow 1 - \alpha$ and $\Pr(\text{select } M_i) \rightarrow 0$ for $i < r$, where r is the true cointegration rank and α is the size of each test in the sequence. These properties do not apply to the testing-down procedure, because the limit distribution of $LR(j - 1|j)$ depends on which submodel M_{j-h} contains the DGP, $h = 1, 2, 3, \dots, j - 1$. Hence one sees that, in a nonstationary setting, the testing-up procedure is recommended, contrary to the common belief (i). Obviously, letting α decrease to zero as sample size increases generates a consistent selection criterion for cointegration rank.

Regarding issue (ii), it is important to note that hypothesis testing can be defined and used also for comparing misspecified models that do not contain the DGP, as shown by Vuong (1989) for independent and identically distributed (i.i.d.) observations. Extension of these results to dependent data and more than a pair of models appears to be a fruitful line of future research.

A final word of caution concerns the distribution of the estimators of the selected model, as shown in recent work by Leeb and Pötscher (2003).

2.3. Multiple Comparisons

Consider finally the case when model selection is the main goal of inference, i.e., one wishes to confront several economic theories with the same set of data. In this context MC procedures may find wider applications in econometrics; see Hsu (1996) for an introduction to MC.

As a simple example, consider the usual regression setup with two regressors,

$$y_i = \sum_{j=1}^2 \beta_j x_{ij} + u_i.$$

Assume also that several competing theories exist on the explanation of y . Theory 1 predicts that $\beta_1 > 0$ and $\beta_2 < 0$, and theory 2 predicts that $\beta_1 = \beta_2$.

Of course one can compute the univariate t tests $t_{\beta_1=0}$, $t_{\beta_2=0}$, and $t_{\beta_1=\beta_2}$ or the corresponding confidence intervals. One can then combine these separate inferences, using probability inequalities such as Bonferroni's (this is called "deduced inference" in Hsu, 1996).

However, in principle, one would like to construct simultaneous confidence intervals for β_1 , β_2 , and $\beta_1 - \beta_2$ from the start, with overall coverage probability not inferior to 95%. This is called "direct inference" in Hsu (1996). Note that the direct method can exploit the fact that we wish to have three confidence intervals for a bivariate distribution of $\hat{\beta} := (\hat{\beta}_1 : \hat{\beta}_2)'$, where a hat indicates the regression estimator.

One possible simple direct inference method is Scheffé's procedure, which states

$$\Pr(c'\hat{\beta} = c'\beta \mp (s^2 c'(X'X)^{-1} ckF_{1-\alpha})^{1/2}, \forall c \in \mathcal{C} \subset \mathbb{R}^k) \geq 1 - \alpha,$$

where $F_{1-\alpha}$ is the $1 - \alpha$ quantile of an $F(k, n - k)$ distribution, and k is the number of regressors, in the example equal to 2. One can form simultaneous confidence intervals for any choice of $c \in \mathbb{R}^k$ using $(s^2 c'(X'X)^{-1} c k F_{1-\alpha})^{1/2}$ as standard error. In our example we can choose c equal to $c_1 := (1:0)'$, $c_2 := (0:1)'$, $c_3 := (1:-1)'$, obtaining three confidence intervals I_1, I_2, I_3 . If the lower endpoint of I_1 is greater than 0 one can assert $\beta_1 > 0$; similarly if the upper endpoint of I_2 is negative, one can assert $\beta_2 < 0$. If finally I_3 contains 0, then one can assert $\beta_1 = \beta_2$. All these statements have a joint error rate bounded by 5%.

Note that, if some new economic theory comes along—theory 3, say, which predicts $\beta_1 + 2\beta_2 = 1$ —one does not have to rework the confidence level of the confidence intervals I_1, I_2, I_3 from the start, but just needs to construct one more confidence interval I_4 for $\beta_1 + 2\beta_2 - 1$.

This feature is both a strength and a weakness of Scheffé's procedure. On one side it gives great flexibility to add any other confidence interval; on the other hand a price is paid in accommodating all these simultaneous confidence intervals. The price is that some possibly less comprehensive procedure gives tighter bounds for each confidence interval; see Hsu (1996).

MCs are widely used in analysis of variance (ANOVA) applications and in regression models; for the application of Scheffé-like procedures in nonlinear regression see, e.g., Johansen and Johnstone (1990). It is possible that many of the procedures within PcGets (see Granger and Hendry, 2005), could be (re)discussed or reorganized in this MC perspective.

3. LEARNING

There appears to be an increasing decoupling among the fields of economic theory (A), econometric theory (B), and applications (C); this is how Mirowski (1995) describes the current age dominated by "critical post-modernism." Separately, the three groups (A, B, C) "maintain a life of their own, through their own journals, their own pedagogy, their forms of tacit knowledge" (Mirowski, 1995, p. 29). Although some degree of autonomy appears constructive (certain "facts" survive changes in economic-theory regimes), a complete separation appears to be contrary to the foundations of all three fields and especially of econometrics (see Frisch, 1933).

One direction that may foster closer cooperation among the three areas is given by the literature on learning, which challenges the notion of rational expectations. Several papers have recently appeared in this line; see, e.g., Guidolin and Timmermann (2003) and references therein.

In these economic models, agents form expectations as an econometrician observing data. Advances in AI could thus help produce better models, e.g., of investors' behavior in financial markets. This may hopefully lead to a better understanding of the actual mechanics of observed economic behavior (C) through the explicit incorporation of the econometric model uncertainty (B)

into the economic model (A). It is not hard to foresee a high reward for teamwork in this area in the future.

4. CONCLUSIONS

In this note I have tried to distinguish between AI and CAD. Both are valuable: AI is better employed in industrial applications and finance, whereas CAD is needed if one wishes to communicate inference, which is at the basis of building economic science.

Information criteria in model selection give many interesting lines of future research. At present there are many IC, and it is difficult to select any single one. Moreover, for non-Bayesians, IC do not provide measures of strength of inference.

Hypothesis testing is still a resource for econometricians, which should be more fully exploited also in the selection of (possibly misspecified) models. Multiple comparisons methods should also be more fully explored in econometrics.

As always, the future of econometrics lies in working more closely with economists; embedding AI rules for investors in economic models of financial markets appears a promising area of development for the economics of learning.

REFERENCES

- Frisch, R. (1933) Editor's note, *Econometrica* 1, 1–4.
- Granger, C.W.J. & D.F. Hendry (2005) A dialogue concerning a new instrument for econometric modelling. *Econometric Theory* (this issue).
- Granger, C.W.J., M.L. King, & H. White (1995) Comments on testing economic theories and the use of model selection criteria. *Journal of Econometrics* 67, 183–187.
- Guidolin, M. & A. Timmermann (2003) Option prices and implied volatility dynamics under Bayesian learning. *Journal of Economic Dynamics and Control* 27, 717–769.
- Hsu, J.C. (1996) *Multiple Comparisons, Theory and Methods*. Chapman and Hall/CRC Press.
- Johansen, S. (1992) Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* 54, 383–397.
- Johansen, S. & I.M. Johnstone (1990) Hotelling's theorem on the volume of tubes: Some illustrations in simultaneous inference and data analysis. *Annals of Statistics* 18, 652–684.
- Kim, J., N. De Marchi, & M.S. Morgan (1995) Empirical model particularities and belief in the natural rate hypothesis. *Journal of Econometrics* 67, 81–102.
- Leeb, H. & B.M. Pötscher (2003) Can one estimate the conditional distribution of post-model-selection estimators? WP available at <http://www.univie.ac.at/seam/Poetscher/poetscher-pub.html>.
- Mirowski, P. (1995) Three ways to think about testing in econometrics. *Journal of Econometrics* 67, 25–46.
- Phillips, P.C.B. (1996) Econometric model determination. *Econometrica* 64, 763–812.
- Pötscher, B.M. (1991) Effects of model selection on inference. *Econometric Theory* 7, 163–185.
- Rao, C.R. & Y. Wu (2001) On model selection (with discussion). In P. Lahiri (ed.) *Model Selection*, IMS Lecture Notes, vol. 38. Institute of Mathematical Statistics.
- Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.