THE SUCCESS OF ECONOMETRICS **

BY

JAN R. MAGNUS*

Summary

THE SUCCESS OF ECONOMETRICS

Econometric theory has achieved much. Still, there is a feeling in the profession that we are not providing the applied economist with the tools that he needs. In this paper I attempt to highlight the things that have occurred to me as being wrong or strange. The link between the items I discuss is the necessity of a focus. I discuss what an econometrician actually does, the relationship between econometrics and physics and econometrics and mathematical statistics, the gap between theory and practice, top-down and bottom-up methodology, weak links, aggregation and hierarchy, and how to take account of other studies. I conclude with an example from the theory of estimation under model uncertainty.

Key words: Econometrics, methodology, top-down, bottom-up, model uncertainty, weak links

1 INTRODUCTION

In this paper I attempt to summarize my thoughts on what might be called the 'pathology' of econometrics. Econometric theory has achieved much, but there is unmistakably a certain amount of gloom among the profession. This gloom, I believe, is the result of the fact that we are not really doing our job. That is, we are not providing the applied economist with the tools that he needs. It is not that we are not providing tools. We are working hard and produce more and more sophisticated tools all the time. They are just not the tools that are required.

More than a century ago, W.S. Jevons wrote:[1]

> The deductive science of Economy must be verified and rendered useful by the purely inductive science of Statistics. Theory must be invested with the reality and life of fact. But the difficulties of this union are immensely great. (Jevons (1871), p.26)

* CentER for Economic Research, Tilburg University, the Netherlands, magnus@kub.nl.

** This paper is a revised and adapted version of my inaugural lecture (in Dutch) at Tilburg University, 25 October 1996. I am grateful to Hugo Keuzenkamp and to the editor and referees of *De Economist* for their constructive comments.

1 Most of the historical references in this paper are taken from Morgan (1990).

De Economist **147**, 55–71, 1999.
© 1999 *Kluwer Academic Publishers. Printed in the Netherlands.*

Immensely great indeed. One difficulty is: how simple should my model be? Tinbergen (1937, p.8), the master of combining simplicity and flexibility, wrote: 'I must stress the necessity for simplification,' while Haavelmo maintained that simple models were legitimate, because if the 'simplified theory also covers the facts, the discovery is an addition to our knowledge' (Haavelmo (1943), p.15). Einstein's attitude was: 'As simple as possible, but not simpler.'

In this paper I shall list some of the things that have occurred to me as being wrong or strange. The link between the items I discuss is the necessity of a *focus*. Of course, I am not the first to criticize the econometric approach. Keynes (1939) and Friedman (1940) criticized Tinbergen (1939) and their points are still valid today. According to Keynes (1939, p.559), Tinbergen had not explained 'the conditions which the economic material must satisfy if the application of the method to it is to be fruitful.' Keynes felt that econometrics could only be applied in a world where there is a correct and complete list of causes, where all the causal factors are measurable, where these factors are independent, where the relationship is linear, and where problems such as time lags and trend factors are adequately dealt with.[2]

I also mention Leamer (1978, 1983), who is concerned with the fragility of the estimates, and Sims (1980), who questions the validity of the economic and statistical assumptions underlying most macromodels.

2 TESTING A TENNIS HYPOTHESIS

Let me begin by discussing an example, from the relatively new field of 'sport statistics,' which shows how difficult it is to translate a seemingly unambiguous question into a testable statistical hypothesis. Tennis commentators often claim that the player who begins to serve in a set has a (psychological) advantage. It is true that most players, when winning the toss, select to serve. Is this a wise strategy?

If there is an advantage, it would be that the player who serves in the first game is usually one game ahead and that this would create extra stress for the other player. In Magnus and Klaassen (1999) we investigated this and some other hypotheses based on data from Wimbledon and we found that 48.2% of the sets played in the men's singles are won by the player who begins to serve in the set and 50.1% in the ladies' singles. (The standard errors of the two estimates are 1.6% and 2.2%, respectively.) Therefore, neither of the two estimates is significantly different from 50%, but, if anything, starting to serve would appear to be a disadvantage rather than an advantage! Closer inspection of the data shows that our finding (starting to serve is a disadvantage rather than an advantage) seems to be true in every set, except the first. In the first set the probability that the

2 See Morgan (1990, section 4.4) for an excellent review of the Keynes-Tinbergen debate.

player who starts serving wins the set is estimated as 55.4% in the men's singles and 56.6% in the ladies' singles. We conclude therefore that in the first set the player who begins to serve has a larger probability of winning the set, but that in the other sets the player who begins to serve has a smaller probability of winning the set. We have now answered a question, but have we answered the right question? Suppose we watch a second set of a match played at Wimbledon. Then it is true that the player who starts to serve has a smaller chance to win this set than his or her opponent. But the reason is *not* that there is a particular advantage or disadvantage in starting to serve. The reason is that, on average, the weaker of the two players will begin to serve in the second set. This can be seen as follows. First, it is much more likely that the last game of the first set is won by a player holding service than by a player breaking service. Secondly, one expects the stronger player to win the first set. Hence, we expect the stronger of the two players to win the first set on his or her own service. This does not always happen, but it happens more often than not. Hence, usually, the weaker player begins to serve in the second set and clearly has a smaller chance to win that set, not because he or she begins to serve, but because he or she is the weaker player.

Where did we go wrong? Our mistake took place in the translation from the verbal question into the statistical hypothesis. We analysed the statistical question 'Is the percentage of sets won by the player who starts to serve higher than 50%?'. Instead we should have analysed the question 'If, at the beginning of a set, you have the choice to begin to serve or receive, is there an advantage to choose to serve?'. The latter question, which is the correct translation of the hypothesis, involves a conditional rather than an unconditional analysis. The answer now is that there is neither advantage nor disadvantage in serving first in a set. Only in the first set is there an advantage in serving first. And this is because the number of breaks in the very first game of the match is extremely small. This gives the server in the first game a small but significant advantage in winning the set.

This non-economic, but simple and highly structured example represents one of the key activities of an econometrician: testing. However simple, the example shows that asking the right question is a non-trivial and difficult aspect of research. It is an exercise in focusing. The example requires a conditional analysis which is typical for all sciences.

Real examples, from economics, are much more difficult. Keynes advocated a modest role for econometrics. He believed that econometrics – the method of multiple correlation, as it was then called – is

> a means of giving quantitative precision to what, in qualitative terms, we know already as the result of a complete theoretical analysis.
>
> (Keynes (1939), p.560)

Most econometricians today are more ambitious. They believe that the main objective of applied econometrics is the confrontation of economic theories with observable phenomena. This involves *theory testing*, for example testing monetarism or rational consumer behaviour. The econometrician's task would be to find out whether a particular economic theory is true or not, using economic data and statistical tools. Nobody would say that this is easy. But is it possible? This question is discussed in Keuzenkamp and Magnus (1995). At the end of our paper we invited the readers to name a published paper that contains a test which, in their opinion, significantly changed the way economists think about some economic proposition. Such a paper, if it existed, would be an example of a successful theory test. The most convincing contribution, we promised, would be awarded with a one week visit to CentER for Economic Research, all expenses paid. What happened? One (Dutch) colleague called me up and asked whether he could participate without having to accept the prize. I replied that he could, but he did not participate. Nobody else responded. Such is the state of current econometrics.

3 ECONOMETRICS, PHYSICS AND MATHEMATICAL STATISTICS

Individuals, households and firms behave so irrationally and their behaviour in groups is so little understood that it is hard to think of an economic law with any claim to universality. This is a strong statement. If the statement is true, this is unfortunate, not only for its own sake, but also because of its consequences. Let me briefly discuss one consequence of a universal law. The example is a very famous one (see Feynman (1965)), taken from physics, a discipline where everything is easier than in applied econometrics, and it shows that when one law is right it can be used to find another one.

If we observe the moons of Jupiter over a long period of time, we discover that the moons are sometimes eight minutes ahead of time and sometimes eight minutes behind time where the time is calculated according to Newton's laws. We also discover that the moons are ahead of time when Jupiter is close to the earth and behind when Jupiter is far away. If we believe Newton, then we must conclude that it takes light some time to travel from the moons of Jupiter to the earth and what we are looking at when we see the moons is not how they are now but how they were the time ago it took the light to get here. Olaus Rømer thus demonstrated in 1675 that light has a finite speed and one year later he estimated that speed at 214,300 km/s, a remarkable achievement, only about 30% too low.

Nothing of such sweeping beauty would ever be possible in econometrics. This is because people, firms, organizations, and their interactions at various levels of aggregation are so much richer and more interesting than planets and therefore, inevitably, much more difficult to model and predict.

Data in econometrics are almost always non-experimental. Wouldn't it be nice if we could double the price of sugar, keeping all other prices the same, and see how the consumer reacts? But economists cannot experiment in this way. Everything moves together. The data that we have to work with are not the result of a controlled experiment. They are ‘non-experimental.’

In physics, chemistry, biology and medicine we can do controlled experiments, but not in economics. (Astronomical data are also non-experimental: we cannot change the orbit of Mars just to see how it affects the orbit of the earth.) This should have a very serious effect on econometric theory. The traditional tools from mathematical statistics – the theory of estimation and hypothesis testing – were designed for the experimental disciplines and not for economics. These tools are therefore not immediately applicable in econometrics.

Related to the non-experimental nature of the data is the fact that econometricians cannot as a rule obtain more data than they already have, certainly in time-series studies. Applied physicists and other experimental scientists can do what the statistical textbooks precribe. They have a theory and collect data, then they form a new theory based on these data, then they throw away the old data and collect new data, then they test the new theory with the new data, and so on. But econometricians cannot do this. If they throw away their old data they have nothing left. In traditional mathematical statistics hypothesis testing and estimation are two separate activities, treated in different chapters or different volumes. The applied statistician either tests a hypothesis or estimates some parameter, but not both at the same time. The econometrician on the other hand is forced to perform both activities simultaneously. Koopmans (1949) suggested that a completely new theory of inference was required to make this possible, but no such theory has been developed. In section 7 I discuss a possible beginning of such a new theory, as it has been developed by Jim Durbin and myself.

4 THEORY AND PRACTICE

There is a gap between theory and practice in econometrics which, I believe, is larger than in physics, medicine and oher disciplines. When Ed Leamer was a graduate student at the University of Michigan, some thirty years ago, there was a very active group building an econometric model of the United States. As it happened, the econometric modelling was done in the basement of the building and the econometric theory courses were taught on the top floor (the third).

> I was perplexed by the fact that the same language was used in both places. Even more amazing was the transmogrification of particular individuals who wantonly sinned in the basement and metamorphosed into the highest of high priests as they transcended to the third floor.
>
> (Leamer (1978), p. vi)

Thirty years later the situation has not changed. Of course, there are some economists and econometricians who find the right balance between what can be done theoretically and what can be done in practice. They know the data, they know economic theory, they know the relevant econometric tools, and they are able to mix these ingredients into a nutritious and tasty meal. But this is a small minority. The large majority either belongs to one camp or the other. My worry as an econometric theorist is not that there is tension between us (the theorists) and them (the applied economists). On the contrary, such tension can be healthy and inspiring. My worry is rather the lack of tension. There are two camps, a gap between them, and little communication.

This theory-practice gap has some rather interesting consequences. One is that applied economists feel the need to test, because they once attended a course called 'econometric theory' and they want to use their skills. But seldom can they explain why they test a certain hypothesis, say homogeneity or convexity. If the hypothesis is rejected – and this is usually the case – they see this as evidence of misspecification. So, why perform a test if the logical consequences of the test are being ignored? One needs to think about the consequences of a test before performing it. Again, this is an example of focusing. But it is not customary in econometric practice.

There is a story, well-know in the profession, which illustrates this gap between theory and practice.

> It is night. A man called P is walking on the street and sees another man, called T, searching for something under a lamppost. 'What are you looking for?' P asks. 'I have lost my keys,' says T. 'Where did you lose them?' P asks again. 'Over there,' says T, pointing some fifty yards further on. 'Then, why are you looking here?' P asks, evidently a practical man. 'It's too dark over there,' T replies and continues to search under the lamppost.

The gap between theory and practice is a problem which should be taken seriously before econometricians lose their credibility. One possible starting point is to start at the other end, not from the theory end but from the applied end. Suppose, for example, we would consider one data set, and ask a number of well-known econometricians to answer specific economic questions by analysing this data set. No other data can be used, but the researcher is free to use his own methods. Such a controlled experiment would, perhaps, allow us to see how sensitive the conclusions are to the methodology used. In fact, this experiment has recently taken place, for the first time in econometrics. The results, reported in Magnus and Morgan (1997, 1999) provide many new insights into methodology and tacit knowledge, but of course not an unambiguous solution.

5 ECONOMETRIC METHODOLOGY

Is there such a thing as an econometric method? Not, of course, a method which takes the innocent researcher by the hand and leads him or her from A to B, then to C and so on. Such a method is non-existent and this is very fortunate, because it would make all of us econometricians jobless. No, I mean something much more humble. Is there any consensus about how to do applied econometrics? When you look at the papers published in the respectable journals, it would seem that there is, because most of theme have the same structure: introduction, what other people have done, the model, description of the data, some problems of econometrics and how the author has solved them, empirical results, conclusion. So there is some consensus about how to *report* empirical work, but this has little to do with how the research was actually *done*. Note, by the way, that these papers never contain a section called 'logbook' where the author tells us how the research was done, in which order, what mistakes he or she made, *et cetera*.

When I was a student at the Institute of Actuarial Science and Econometrics at the University of Amsterdam we were taught (at least that is how I understood it) that you should build a model, collect data, select the appropriate method of estimation and estimate the model. Then we could proceed in several directions: estimate certain functions of the parameters, such as elasticities, test hypotheses of interest, make forecasts, or make policy recommendations. This is a nice method, but it does not work. It is far too ambitious. In economics, unlike physics, there are no models which remain valid in all directions. The best we can hope for is that a model is valid locally. This implies that the model should depend on the central question which the researcher wishes to answer. I call this the *focus* of the research. In my view, choosing your focus is the only sensible way to start. Everything else – your model, the data that you need, your estimation method – depends on it. Now, this may seem obvious, but it is not obvious to most econometricians. I shall discuss the concept of a focus and how it can be used later on.

Econometricians differ of course on how to do applied econometrics and several debates have recently taken place. Both the 1985 World Congress of the Econometric Society in Cambridge, Mass. and the 1988 Australasian Meetings in Canberra contained a major methodological panel discussion; see also Pagan (1987). In listening to these debates, there appears to be one thing – and one thing only – in which we as econometricians all believe. This is the top-down approach.[3] In the top-down approach we are told to start off with a large model, lots of variables and then successively test the relevance of the variables. If a variable does not turn out to be relevant (after a statistical test), we delete it. Proceeding in this way would produce a smaller model with only the relevant variables in it. In theory this sounds fine and it has certain theoretical niceties. But there are problems. The first problem is that it does not work. If you try to estimate such a large model, which has everything in it that you can think of, you get nonsensical results. Everyone who has done empirical econometric work

3 I realize, of course, that not *all* econometricians follow the top-down approach.

knows this. The second problem is that you cannot really discover anything new and interesting in this way, because the interesting bit is the construction of the top model and we are not told how this is done. Therefore no applied economist proceeds in this way. Instead they follow the bottom-up approach. In the bottom-up approach one starts with a simple model and builds up from there. This is, in fact, how scientists in other disciplines work.

So, we see that there is some theory on the top-down approach, which nobody uses, but that there is no theory on the bottom-up approach, which is used in practice. The applied economist works bottom-up, but there is no theory to tell him how to proceed from small to large, for example how to select a new variable if he feels that the small number of variables in the simplest model is not sufficient. I am not saying that this is an easy problem for the theoretical econometrician, but at least it would be something that applied workers could use.

A sad consequence of the clash between the top-down theory and the bottom-up practice is the presentation of the results. Here the applied economist has a problem, because he has obtained the result using the bottom-up approach, but the journal tradition prescribes that a serious article should follow the top-down approach. The practical solution is as follows. First the applied economist plays with the data, starts off with a small model, adding ingredients to it, until he is happy. This is bottom-up. In order to satisfy the editor of the journal, he then goes one step further and adds some ingredients to the model that he really does not believe should be in there at all. This extended model is the one he then presents at the beginning of paper. The researcher, following the top-down prescriptions, now tests whether the things he has just put in can be deleted and arrives at the model he wants to work with. This may sound like a swindle. I think it is.

6 FIVE SPECIFIC PROBLEMS

6.1 *Weak Links*

Let us consider briefly what a theory of the bottom-up approach would entail, remembering that no such theory yet exists. It seems to me that one aspect of this new theory should be that it helps us decide what the weak links in a model are. An empirical piece of work has many aspects. There are the data, the economic theory, the econometric method of estimation, and other aspects. The researcher starts with a simple model, a simple method of estimation and some readily available data. He gets some results but is not quite satisfied. Then what? Do we need to extend the model, improve the estimation technique or do we need more (or better) data? The last option (improve the data) is the most humble and often the most sensible option, but is seldom chosen by econometricians. Instead we like to concentrate on the model and the method of estimation. Far too often one sees subtle improvements in the technique of estimation (e.g. full information maximum likelihood instead of instrumental variables, nonparametrics instead of simple probit) even when the data contain a lot of noise and it would have been much better to spend some time getting solid data. Clearly there should be a balance between the different ingredients of the empirical study and the weakest link should be discovered and improved. Now, is there a test which tells us what the weak link is? No, such a test does not exist. Of course it does not exist, because we have already seen that there is no bottom-up theory. Hence, a 'weak link' test, which is an aspect of bottom-up, does not exist either. Such a test would be more general than a specification search, because it would include the possibility that the model is reasonable, but the data have too much noise or the estimation technique fails.

Is a 'weak link' test conceivable? I think it is. One way to develop the test would be to think of a collection of sensitivity tests in various directions. The main difficulty then is to find the proper weights for these directions. No solution is yet in sight for this problem, but in addressing it we are at least searching for the keys where they were lost and not where we wished they were lost.

6.2 *Aggregation*

There is more. Related to the problem of 'weak links' is the question of aggregation. Many researchers are interested in estimates of functions of parameters in macro relationships (for example, price elasticities in a macro setting). Obviously one can obtain these from a macro model. But one can also obtain these estimates as a weighted average of estimates from micro relationships. At first sight the latter approach would seem preferable, even though it requires many more data. As it turns out, the micro approach is not always preferable, probably because the micro relationships may behave very differently from the macro relationship.[4] But even if the micro approach would seem preferable, it is much more work and more expensive. The relevant question here is whether one can decide, based on macro data and perhaps rudimentary and incomplete micro data, whether it is worthwhile to make the micro data set more complete. I believe such a decision rule (or test) is possible. It would be a useful tool to have, but it does not yet exist.

6.3 *Hierarchy*

Another question, related to the question of aggregation, is the modelling of a hierarchy. This is not a common procedure in applied econometrics. One either

4 I studied this problem within the context of energy demand with Alan Woodland; see Magnus and Woodland (1987, 1990).

uses micro or macro data but seldom both. A hierarchy is a set of levels of aggregation (say, micro and macro, but possibly more), where each level helps to explain certain features of the economic phenomenon under consideration. The importance of estimating hierarchies lies in the realization that aggregation is not linear and that therefore the micro processes will usually behave very differently from the macro process. In economics the modelling of a hierarchy is a rather novel approach and one where econometric theory can play an important part. Using hierarchies we can combine micro and macro theory and obtain better and more robust estimates and forecasts.

6.4 *How to Take Account of Other Studies?*

I now come to a very fundamental point which is related to aggregation and the modelling of hierarchies in the sense that it also concerns the relationship between various pieces of relevant information. Suppose you wish to study the substitution possibilities between capital, labour and energy in Dutch manufacturing, as I did in Magnus (1979). There are two closely related studies, one with Canadian data and one with US data. What do we do with these other studies? Clearly they contain some relevant information, but how can this information be properly incorporated in the Dutch study? A Bayesian approach might be the solution, although this approach has its own difficulties, for example – but not only – the choice of priors. But most of us are non-Bayesians or semi-Bayesians. So, what do we do? The common procedure is as follows. We read these other studies, but since we want our own contribution to be as novel as possible, we do not exactly copy the model or estimation technique of either paper. Instead we try something a little different. The related papers are mentioned in our introduction, but otherwise ignored until we come to the conclusion. Now there are two possibilities. Either our results are close to the previous studies or they are not. If they are, we are happy and tell the reader that clearly our study was a good and solid one, confirming for the Netherlands what had already been found elsewhere. If, on the other hand, our results are not close to the previous studies, we are also happy and tell the reader that the result is no surprise, because our method is different and what the others did is clearly wrong. I do not think this is a very good or honourable procedure. Again, the problem is essentially unsolved, but it is a fundamental problem worth taking seriously.

6.5 *The* t-*ratio*

The famous *t*-ratio plays two roles in econometrics, one proper and one improper. The proper role involves *testing*, namely testing whether a particular coefficient is close (in a statistical sense) to an ex ante given number, usually zero. The coefficient is a coefficient of interest to the researcher, and the researcher wishes to find out from the data whether the coefficient equals zero (for example) or not. There are some problems with this procedure in terms of errors of the first kind versus errors of the second kind, but basically the procedure is sound.

The improper role of the t-ratio involves *modelling*. The difference with the first situation is that the coefficient under investigation is now a nuisance parameter rather than a coefficient of interest. Suppose for example that we consider a standard regression model

$$y = X\beta + \gamma z + u, \tag{1}$$

where y is explained by k regressors in the matrix X and one additional regressor z. We are uncertain whether z should be in the model or not and hence we look at the t-ratio for γ. If this turns out to be 'large' (say, larger than 1.96 in absolute value), then we conclude that we should keep z in the model, but if the t-ratio is 'small' (smaller than 1.96), then we say that the estimator for γ is not significant and we delete z from the model. This use of the t-ratio is commonplace in econometrics, but it makes no sense at all. This is because the focus is wrong. The question is not whether γ is zero or not, but whether including z in the model produces a better estimator of β. This question was already recognized by Frisch who, discussing problems of demand analysis, wrote:

> In this field we need, I believe, a new type of significance analysis, which is not based on mechanical application of standard errors computed according to some more or less plausible statistical mathematical formulae, but is based on a thoroughgoing comparative study of the various possible types of assumptions regarding the economic-theoretical set up, and of the consequences which these assumptions entail for the interpretation of the observational data. (Frisch (1933), p.39)

The question is further investigated in the next section.

7 ESTIMATION UNDER MODEL UNCERTAINTY: AN EXAMPLE

Let me try and explain what I mean with the help of the following story, taken from Magnus (1999).

> On the tiny remote island of I, the I-landers lived mainly on fish. Since the wind around the island varied from day to day, the I-landers had built two boats for their fishermen, the R-boat and the U-boat. The R-boat (R for 'rest') was ideal when there was no wind, the U-boat (U for 'unrest') was ideal in a heavy storm. Each evening after dinner the King announced his weather forecast for the next day upon which one of the two boats was prepared. An incorrect forecast of the weather and hence a

> wrong choice of boat could have serious consequences. All I-landers remembered the recent hurricane, which the King had failed to forecast and where the R-boat capsized, resulting in the death of all fishermen on board. One day a young adventurer A found himself stranded at I. A inspected the two fishing boats and found them well-built for their purpose. He noticed, however, that the extreme weather conditions for which the boats were designed rarely occurred. Most days at the island saw a moderate breeze. A decided to build a boat himself. After several months, his work completed, he proudly presented his new boat to the assembled I-landers. 'How does your boat perform when there is no wind?' asked one of the fishermen. 'Well,' said A, 'you can't expect my boat to do quite so well as your R-boat, which was built for that purpose, but it definitely does better than the U-boat when there is little or no wind.' 'And how does your boat perform in stormy weather?' asked a second fisherman. 'Again,' answered A, 'your U-boat does better in a storm, but my boat performs better than the R-boat. In particular, my boat will not capsize in a storm.' Then the King said: 'If the weather is fair, with a gentle breeze, what performance does your boat have then?' Somewhat embarrassed A replied: 'I must admit that under such conditions my boat performs worse that both the R-boat and the U-boat.' 'Throw this man into the ocean!' cried the King, and A was never heard of again.

We can even draw a picture of the risk of the three boats as a function of the wind force; see Figure 1 (ignore the curve labelled MD's boat for the moment). Let us assume that wind forces below one and above one occur equally often. Wind force one represents the 'gentle breeze' that the King was referring to. If the wind force is smaller than one we should take the R-boat; otherwise the U-boat. The King, of course, is uncertain about the wind force. The problem with A's boat is that it never has a smaller risk than both the R-boat and U-boat and that around one (which is the most important area) it has a higher risk than the other two boats. A's boat is therefore not acceptable. Can we construct a boat which has a smaller risk than both the R-boat and the U-boat for wind forces around one and which performs well in extreme conditions too? It turns out that we can. The risk of MD's boat (modestly named after its designers: Magnus and Durbin) has these properties.

The story corresponds exactly to a situation which occurs in econometrics all the time, namely how to properly estimate the parameters of interest β in the linear regression model (1), where the explanatory variables in X are regarded as belonging to the equation according to some theory, and can be thought of as the minimum set of variables required to explain y. The explanatory variable z, however, is only included because the researcher believes it might lead to 'better' estimates of β. The focus of our analysis is the estimation of one or several lin-

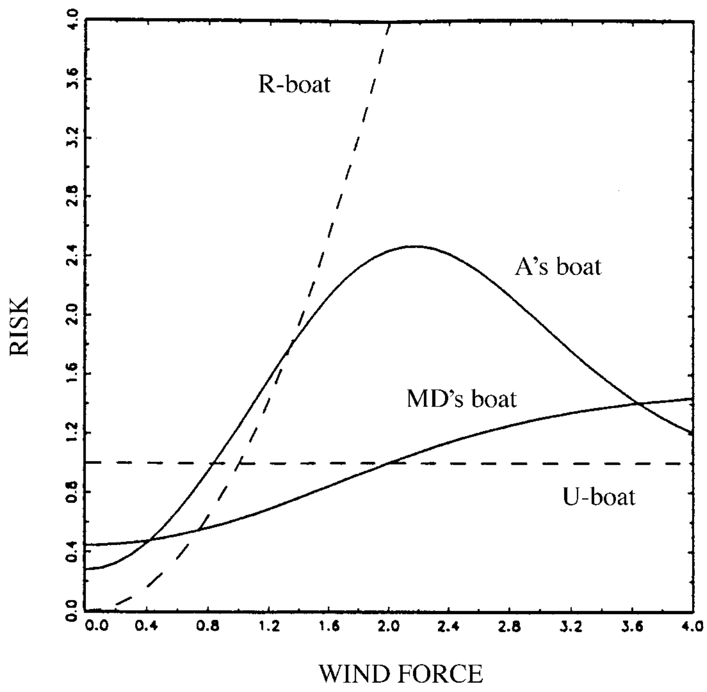

Figure 1 – Risk of the four boats

ear combinations of the parameter β, while we are not interested in γ. This is an old and classical problem in applied statistics.

Consider first the unrestricted least squares estimator for β, denoted by b_u. Its risk (mean squared error) corresponds with the risk of the U-boat in Figure 1. The wind force is the translation of the 'theoretical t-ratio' of γ, that is, γ divided by the square root of the variance of its least squares estimator. Let θ denote this theoretical t-ratio. If θ is large (in absolute value), then z should be included in the model and b_u performs well. But if θ is small (close to zero), then b_u does not perform well. The situation is reversed for the restricted least squares estimator b_r, estimated under the restriction that $\gamma = 0$. Its risk is given by the risk of the R-boat. We see that for large θ its risk is even unbounded.

Neither of these two extremes (b_u or b_r) is satisfactory. The usual procedure is therefore to test whether z should be in the model by a t-test on γ. If the t-ratio (in absolute value) is smaller than, say, 1.96, the econometrician chooses the restricted estimator b_r; otherwise the unrestricted estimator b_u. The resulting estimator is the so-called 5% pretest estimator. Its properties are neither those of b_r, nor those of b_u. The estimator corresponds to A's boat in the story and its risk is given in Figure 1. The pretest estimator makes a laudable attempt to combine estimation and testing, but unfortunately not a very happy one. In particular, the risk of the pretest estimator is higher than the risk of both the restricted and the unrestricted estimator in the area where it matters most (around 1)! In spite of this, the pretest estimator is the most commonly used estimator in econometrics.

Is it possible to combine estimation and testing in a better way? Suppose we estimate β as a weighted average of the unrestricted and the restricted estimator, that is,

$$b = \lambda b_u + (1 - \lambda) b_r, \tag{2}$$

where λ is a function of the t-ratio of γ; see Magnus and Durbin (1999) and Magnus (1998). The pretest estimator is a very simple example of such an estimator. The problem is how to choose the weight function λ. We derived an estimator (the 'ideal' Laplace estimator) which is near-optimal and also has an attractive Bayesian interpretation. This is our boat, MD's boat, in Figure 1. We believe that the 'ideal' Laplace estimator (MD's boat) could be the beginning of a new theory of inference based on the idea that we require in econometrics a theory capable of hypothesis testing and estimation at the same time.

Apart from the conclusion that combining the theory of estimation and the theory of hypothesis testing is possible and that we have a situation here where compromising is good rather than bad, we see how crucial it is to choose a focus. Our focus was the estimation of β. Had we chosen a different focus, for example to forecast y, then we would have obtained a different estimator for β. In particular, our focus was *not* to select the best model. The correct model is almost certainly the unrestricted model, because γ is almost certainly not equal to zero. But the correct model does not necessarily produce the best estimates! Misspecification is only bad if specification is the focus, but not necessarily if estimation is the focus.

8 CONCLUSION

In this paper I have attempted to list some of the things that have occurred to me as being wrong or strange in econometric theory and practice. I could have added many other things. Let me briefly mention four of these.

The neglected role of paradox. Paradoxes can provide important new insights: 'misspecification can yield better estimates of the parameters of interest,' 'the further ahead you forecast the more precise is your forecast' (Hoque, Magnus, and Pesaran (1988) and Magnus and Pesaran (1989)), 'research assistants should not be too good' (Rothenberg (1985)).

Disasters and structural breaks. I think it is beyond econometrics to predict a structural break such as the fall of the Berlin Wall in the autumn of 1989. But at least we should have a theory which can be used to tell us how our estimates and forecasts change after a structural break. Such a theory does not exist, but it is not impossible. Lucas (1976) in his famous 'critique' questioned whether the structural parameters in economic relationships remain constant under a policy intervention. The question is not whether the parameters change (of course they do), but whether this has any serious impact on the problem under discussion and, if so, what this impact is.

Data formation. Why do we not assume responsibility for the data we use? If someone attacks us by pointing out a flaw in the data, we respond that this is not our fault, but the fault of whoever supplied the data. No other discipline would accept this lack of reponsibility. Also, the data are not value-free, but often collected according to some economic theory. For example the filling in of missing observations is often done using the most relevant economic theory. Thus in testing an economic hypothesis one jointly tests the data and the theory; see Griliches (1986) for an excellent exposition and Morgenstern (1950) for some early warnings.

Linearity. What is the role of linearity beyond the obvious idea that a linear model is a first-order approximation to a smooth nonlinear model and hence provides a good approximation under certain conditions? Many of the concepts in statistics are linear. The normal distribution is intimately connected to the linearity fallacy (every linear combination of normal variates is normal) and so is the dangerous concept of unbiasedness.

Fortunately, there is still a lot of work to be done. Econometrics has had wonderful successes and econometric theory has developed with terrific speed and depth. Nevertheless, we have not been searching for the keys where we expect them to be. So, we find all sorts of interesting things under the lamppost: a coin, a piece of string, and so on, things that we proudly report on in our research papers, but we do not find the keys. This is depressing. There is only one solution, namely to move the lamppost. That done, econometricians can continue to make important contributions and eventually, perhaps, become respectable scientists. As Stigler wrote many years ago:

> The inconclusiveness of the present statistical demand curves should not be taken as an excuse for lapsing back into the use of 'common sense' and 'intuition'. G.J. Stigler (1939), p.481)

REFERENCES

Feynman, R.P. (1965), *The Character of Physical Law*, London, Penguin Science.

Friedman, M. (1940), 'Review of Jan Tinbergen. Statistical Testing of Business Cycle Theories, II: Business Cycles in the United States of America', *American Economic Review*, 30, pp. 657-661.

Frisch, R. (1933), 'Pitfalls in the Statistical Construction of Demand and Supply Curves', *Veröffentlichungen der Frankfurter Gesellschaft für Konjunktur Forschung*, New Series, V, Leipzig, Hans Buske.

Griliches, Z. (1986), 'Economic Data Issues', in: Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, 3, Amsterdam, North-Holland, chapter 25.

Haavelmo, T. (1943), 'Statistical Testing of Business-Cycle Theories', *Review of Economics and Statistics*, 25, pp. 13-18.

Hoque, A., J.R. Magnus, and B. Pesaran (1988), 'The Exact Multi-Period Mean-Square Forecast Error for the First-Order Autogressive Model', *Journal of Econometrics*, 39, pp. 327-346.

Jevons, W.S. (1871), *The Theory of Political Economy*, London, Macmillan.

Keuzenkamp, H.A. and J.R. Magnus (1995), 'On Tests and Significance in Econometrics', *Journal of Econometrics*, 67, pp. 5-24.

Keynes, J.M. (1939), 'Professor Tinbergen's Method', *Economic Journal*, 49, pp. 558-568.

Koopmans, T. (1949), 'Identification Problems in Economic Model Construction', *Econometrica*, 17, pp. 125-144.

Leamer, E.E. (1978), *Specification Searches*, New York, John Wiley.

Leamer, E.E. (1983), 'Let's Take the Con out of Econometrics', *American Economic Review*, 73, pp. 31-43.

Lucas, R.E., Jr. (1976), 'Econometric Policy Evaluation: A Critique', in: K. Brunner and A.H. Meltzer (eds.), *The Phillips Curve and the Labor Market*, Carnegie Rochester Conference Series, 1, Amsterdam, North-Holland, 19-46.

Magnus, J.R. (1979), 'Substitution Between Energy and Non-Energy Inputs in the Netherlands, 1950-1976', *International Economic Review*, 20, pp. 465-484.

Magnus, J.R. (1998), 'Estimation of the Mean of a Univariate Normal Distribution with Known Variance', *Econometric Theory*, forthcoming.

Magnus, J.R. (1999), 'The Traditional Pretest Estimator', *Theory of Probability and Its Applications*, 44, forthcoming.

Magnus, J.R. and J. Durbin (1999), 'Estimation of Regression Coefficients of Interest When Other Regression Coefficients Are of No Interest', *Econometrica*, 67, forthcoming.

Magnus, J.R. and F.J.G.M. Klaassen (1999), 'On the Advantage of Serving First in a Tennis Set: Four Years at Wimbledon', *The Statistician*, 48, forthcoming.

Magnus, J.R. and M.S. Morgan (1997), 'The Experiment in Applied Econometrics', *Journal of Applied Econometrics*, 12, Supplement.

Magnus, J.R. and M.S. Morgan (1999), *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, Chichester/New York, John Wiley and Sons.

Magnus, J.R. and B. Pesaran (1989), 'The Exact Multi-Period Mean-Square Forecast Error for the First-Order Autoregressive Model With an Intercept', *Journal of Econometrics*, 42, pp. 157-179.

Magnus, J.R. and A.D. Woodland (1987), 'Inter-fuel Substitution in Dutch Manufacturing', *Applied Economics*, 19, pp. 1639-1664.

Magnus, J.R. and A.D. Woodland (1990), 'Separability and Aggregation', *Economica*, 57, pp. 239-247.

Morgan, M.S. (1990), *The History of Econometric Ideas*, Cambridge, Cambridge University Press.

Morgenstern, O. (1950), *On the Accuracy of Economic Observations*, Princeton, Princeton University Press.

Pagan, A.R. (1987), 'Three Econometric Methodologies: A Critical Appraisal', *Journal of Economic Surveys*, 1, pp. 3-24.

Rothenberg, T.J. (1985), *Incredible Structural Inference*, Department of Economics, The University of California at Berkeley.

Sims, C. (1980), 'Macroeconomics and Reality', *Econometrica*, 48, pp. 1-49.

Stigler, G.J. (1939), 'The Limitations of Statistical Demand Curves', *Journal of the American Statistical Association*, 34, pp. 469-481.

Tinbergen, J. (1937), *An Econometric Approach to Business Cycle Problems*, Paris, Hermann and Cie.

Tinbergen, J. (1939), *Statistical Testing of Business Cycle Theories*, 2 volumes, Geneva, League of Nations.