

ORGANIZATION OF THE EXPERIMENT

JAN R. MAGNUS^{a*} AND MARY S. MORGAN^b

^a*CentER for Economic Research, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands*

^b*University of Amsterdam and London School of Economics, Houghton Street, London WC2A 2AE, UK*

SUMMARY

This paper explains the organization of our experiment in applied econometrics, where participating teams had one year (1 July 1995 to 1 August 1996) to answer specific economic questions with a specified data set, but using their own methodology and economic insights. In the end, eight teams completed their task. We give detailed information about the rules and the tasks, we provide the essence of our correspondence with the teams and the assessors, and summarize what happened at the workshop in Tilburg (December 1996) and afterwards. © 1997 John Wiley & Sons, Ltd.

J. Appl. Econ., 12, 467–476 (1997)

No. of Figures: 0. No. of Tables: 0. No. of References: 12.

1. INTRODUCTION

This paper contains a technical description of the organization of the experiment. We had been thinking about an experiment for many years and discussed it with many people. The experiment in its current form was first announced in the May 1995 issue of the *Journal of Applied Econometrics* (Magnus and Morgan, 1995a). We told prospective participants that we had selected one classic paper—Tobin (1950)—and we briefly summarized the purpose, the data, the tasks, and the assessments of the experiment. Then we invited participants to come forward and perform the described tasks within one year. In total, 39 individuals or teams wrote to us. (Four teams who wanted to participate after the deadline were allowed to do so.) The teams were geographically distributed as follows:

	Participants	Completed
USA	14	4
UK	7	1
Netherlands	4	2
Rest of Europe	8	1
Rest of World	6	0

On 1 July 1995 we sent to each participant:

- The *Experiment Information Pack*, see Magnus and Morgan (1995b)
- A reprint of the Tobin (1950) article
- One diskette containing the data.

We anticipated that there would be questions from the participants during the year. In order to sustain complete equality of information, we promised that, if at all appropriate, we would send

* Correspondence to: Jan R. Magnus, CentER for Economic Research, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

the question and our answer to all participants. We also promised that, if we discovered relevant papers, we would inform participants about these. Papers mentioned at the start as possibly useful were Chetty (1968), Maddala (1971) and, in particular, Izan (1980).

In Section 2 we describe the rules of the experiment. Section 3 contains the tasks that we set to the participating teams. In Section 4 we ask the participants to keep a logbook and provide instructions on the report. A summary of our correspondence with the participants by e-mail during the year is given in Section 5. The assessors and their role in the experiment is described in Section 6. Section 7 explains the organization of the workshop in Tilburg. Finally, Section 8 summarizes our activities between the workshop and the publication of this Special Issue.

2. THE RULES¹

We believed that strict rules were necessary for the success of the experiment. Comparability of the results (difficult in the best of worlds) would depend on strict controls of certain aspects of the experiment. We attempted, however, to keep the rules as simple as possible. These were our rules.

Use of the Data

- (a) Only data supplied by Magnus and Morgan (hereafter MM) can be used. However, there is no restriction on the use of economic theory and econometrics.
- (b) The US data is drawn from official published reports and is in the public domain. Please credit the relevant US government department and cite the relevant publications (all references are given in Chapters 7 and 8 of the *Experiment Information Pack*) when you make use of this data. Some of the Dutch data can only be used courtesy of Statistics Netherlands. Each report should contain a note of thanks to Statistics Netherlands.
- (c) Participants are not permitted to provide the data supplied by MM to anybody other than members of their research team.

Time Schedule

A strict time schedule is a necessary condition for the success of this experiment. The date and location of the workshop are fixed and the date that copy will have to be with the *Journal of Applied Econometrics* is fixed as well.

1 July 1995	Information packs have been sent out
1 July 1996	All experiment reports with MM
1 September 1996	Selected reports to assessors
14–17 December 1996	Workshop at CentER, Tilburg University, the Netherlands
1 March 1997	Assessors comments to MM
1 October 1997	Publication of Special Issue of the <i>Journal of Applied Econometrics</i> (Volume 12, Supplement)

The time schedule is holy.

¹ This section is a slightly adjusted version of Chapter 2 of the *Experiment Information Pack*; see Magnus and Morgan (1995b).

Exchange of Information

No information concerning the ongoing experiment should be exchanged between participating experimenters before 1 July 1996.

Reporting²

- (a) The focus of each report must be the tasks set by MM.
- (b) All results must be reproducible.
- (c) A logbook of the process by which the results have been obtained must be kept and a summary logbook must be submitted as one section of the final report.

Publication of the Results

Participants are obliged to submit their reports to MM on or before 1 July 1996. The report shall be considered as a submission to the Special Issue of the *Journal of Applied Econometrics* and will be refereed according to the highest academic standards. The report (or parts of it) should not be published elsewhere until after publication of the Special Issue.

Role of MM

- (a) MM will act as organizers and administrators of the project.
- (b) MM will serve as guest editors of the Special Issue of the *Journal of Applied Econometrics* and have been given complete editorial responsibility.
- (c) MM will not participate as one of the research teams producing a report and aim to be neutral between approaches.
- (d) In cases where the above rules do not apply or where a change of the rules is deemed necessary, the decision by MM is final.

3. THE TASKS³

There were five tasks. We hoped participants would attempt to undertake all five tasks, but we made it clear that this was not necessary if a particular task could not be performed within the framework of the participant's methodology. The tasks were described as follows.

3.1. Measurement

- (a) Using the original Tobin data set for the USA:
 - (i) Estimate the income elasticity of food demand;
 - (ii) Test the 'homogeneity postulate' of the family food demand function;
 - (iii) Comment on the differences between your results and Tobin's.

Note: You may use the data set as used by Tobin (with or without our corrections) and, if you wish, the additional data that Tobin could have used but didn't.

- (b) Using the full set of US data provided, perform the tasks specified under (a).

² See also Section 4—The logbook and report.

³ Slightly adjusted from Chapter 3 of the *Experiment Information Pack*; see Magnus and Morgan (1995b).

3.2. Measurement with Related Information

- (a) Assume now that you have no access to the US data set, only to the Dutch set. However, a colleague of yours (whose work you moderately admire) has recently published a study on the US demand for food in the *Journal of Applied Econometrics* using the US data set. The *results* of this study (estimates, standard errors, tests, predictions) are available to you, but *not* the underlying data. In particular, your American colleague has performed Task 1(b).

You are asked to undertake the tasks specified under 1(a) for the Netherlands using the Dutch data set. In doing so you wish to take account of the American article. You also wish to take advantage of the experience of your American colleague. Upon your request you are given the logbook underlying the American study. How do you take account of the American article and/or the logbook in your Dutch study? If you had completely ignored the American article, would your results have been any different?

- (b) Assume next that both data sets are available to you. Thus you have more information than in Task 1(b) (because you may use the Dutch set) and Task 2(a) (because you can now use the US data set, not just the resulting US estimates). Answer the questions in Task 1(a) again for both the USA and the Netherlands and comment on the difference (if any) between the results of Task 1(b) and Task 2(b) for the USA and between the results of Task 2(a) and Task 2(b) for the Netherlands.

3.3. Forecasting

In 1988 (the last year of the Dutch time series available to you) the research department of an important food manufacturer, operating mainly in the Dutch market, wished to make economic forecasts until the year 2000. You were hired as a consultant. Produce forecasts for the demand for food in the Netherlands as it will develop in the twelve years from 1989 to 2000.

3.4. Policy

Governments and policy makers are now worried about the differential impacts on different population groups of changes in the economy. One aspect of this problem is the relationship between income distribution and aggregate economic performance (or real income performance). Is a flat income distribution good or bad for the economic performance of an economy? Does an improved economic performance change the income distribution? If so, how? The relationship between income distribution and economic performance thus has a causality aspect as well as a quantitative aspect.

- (a) Construct an index of aggregate real income performance for each of the two economies.
- (b) Investigate the relation between income distribution and aggregate real income performance.
- (c) What would be the effects (if any) of tax policies designed to redistribute incomes?

Note: Some references relating to this question are: Blank and Blinder (1986), Blinder and Esaki (1978), Carroll and Summers (1991), Cutler and Katz (1991), and Quah (1994).

3.5. Own Task

Define your own task (within the context of the data sets provided). Describe the problem, your procedures and results.

4. THE LOGBOOK AND REPORT⁴

The Logbook

Most scientific research relies on a large amount of ‘tacit knowledge’ — knowledge which is not part of formal theoretical ideas (either statistical or economic), but which is an essential element in conducting applied scientific research. It is this knowledge, gained from experience, which guides the scientist in deciding what to do next and how to do it. It seems reasonable to assume that this is the case in econometrics as well.

In an attempt to learn more about the process of applied econometrics, and to access the sort of tacit knowledge involved, we asked all participants to join us in an attempt to throw light on this important aspect of econometric research. As a mechanism for keeping track of the process we suggested the ‘logbook’. It is commonplace in other scientific fields to keep lab notebooks. These record the procedures used, the various steps taken as the research progresses, false avenues, interim results, and other details the scientist wishes to keep track of. These records, directly or indirectly, can reveal much about the research process. We asked all our participants to keep such notebooks, which we call ‘logbooks’ and which form an important element of the experiment.

There are no hard and fast rules about what should be in a logbook. Different participating groups might keep note of different aspects. But at a minimum we think it should include notes on data preparation work, the order of work on each task, failed steps as well as successful ones, and so forth. The logbook record ought to be sufficient for the author to reconstruct the path of the author’s work and the reasoning behind it.

We asked that a summary of the logbook be included in the final report in a shortened form (2–3 pages). In this way, other participants (and later the readers of the reports), would be able to learn about the process of applied econometrics.

The Report

We told participating teams that the Special Issue of the *Journal of Applied Econometrics* would include a full description of the experiment including the tasks, data description and sources. These should therefore *not* be repeated in each author’s report. Instead, participants’ reports should focus on how the tasks are undertaken, the outcomes interpreted, and the results assessed according to self-set criteria.

We asked the authors to bear in mind the following guidelines and requests:

- (a) Please keep to a maximum of 25 pages.
- (b) Please include a short summary of your overall approach/methodology (include references to longer explanations where necessary).
- (c) Please outline your own criteria of success.
- (d) Please include a brief, honest summary account (including false moves) of your progress on each task based on your logbook record.
- (e) Please do not change the variable names unnecessarily.
- (f) Please include your interpretation/assessment of your results, including those things you are disappointed with as well as those you are pleased with.
- (g) Please describe the self-set task and its results.
- (h) Please remember that all results should be reproducible.

⁴ Slightly adjusted from Chapter 4 of the *Experiment Information Pack*; see Magnus and Morgan (1995b).

5. E-MAIL MESSAGES SENT TO PARTICIPANTS

During the year we kept in touch with all participants through e-mail. We encouraged them to ask for clarification if necessary, informed them about new participants, asked their advice about possible assessors, reminded them of the deadlines and the rules of the experiment, and corrected some errors. Of these messages, one (19 August 1995) contained some additional data for the 1941 US budgets survey. On 1 May 1996 we provided a corrected data set for Dutch population figures—with our apologies. Because of this error on our part, and because of requests from participants, we postponed the deadline for completion of the experimental reports until 1 August 1996. These two e-mail messages and associated data will be provided in Magnus and Morgan (1998), and will be available as additional information to the data archived with the *Journal of Applied Econometrics*.

Only one e-mail message related to clarification of the experimental design: On 29 August 1995 we e-mailed participants as follows in reply to one participants' comment on Task #2 (measurement with related information):

Comment:

- (a) This question could be interpreted as asking for a hypothetical response to hypothetical stimuli, which (as the participant pointed out) contrasts greatly to the implicit philosophy of everything else, namely real responses to real stimuli.
- (b) The first sentence: 'Assume next that both sets are available to you' might suggest a hypothetical circumstance which was not intended.

Our reply:

- (a) The idea here is to use your own estimation results on the USA. Thus the 'colleague whose work you moderately admire' is yourself. This is slightly schizophrenic but it is the only way this task can be realistically performed;
- (b) The sentence should read: 'Next use both data sets available to you.'

6. THE ASSESSORS

During the summer of 1996 we were able to find eight assessors to read and comment on the reports presented at the workshop. They are (in alphabetical order):

Short name	Assessor	Affiliation
AB	Anton Barten	Catholic University of Leuven, Belgium
JSC	J. S. Cramer	University of Amsterdam, the Netherlands
AK	Arie Kapteyn	CentER, Tilburg University, the Netherlands
MMc	Michael McAleer	University of Western Australia, Australia
HP	Hashem Pesaran	Trinity College, Cambridge, UK
PS	Peter Schmidt	Michigan State University, USA
KWa	Kenneth Wallis	University of Warwick, UK
MWi	Michael Wickens	University of York, UK

We received reports from eight teams who completed the experiment. They are (in the order of presentation at the December workshop):

Short name	Authors	Assessed by
CBS	van Driel/Nadall/Zeelenberg	AK, HP, MWi
Texas	Anderson/Vahid	JSC, PS, HP
Leamer	Leamer	AB, MMc, PS
Maddala	Maddala/Wu/Yin	JSC, AK, MWi
Maastricht	de Crombrughe/Palm/Urbain	AB, MMc, KWa
Finland	Höglund/Jäntti/Knif/Nordberg/Rosenqvist	AB, PS, KWa
Dundee	Song/Liu/Romilly	JSC, KWa, MWi
Tennessee	Bearse/Bozdogan/Schlottmann	AK, MMc, HP

There are thus eight reports and eight assessors. Each report was sent to three assessors and each assessor was asked to assess/referee three reports.

In our letter of 26 August 1996 to the assessors we sent each of them:

- The information pack (54 pages)
- Data diskette (9 original files +5 ASCII files concerning the US data)
- The Tobin article
- E-mail messages sent to participants (4 pages of clarifications/corrections)
- Three reports to assess.

We asked each assessor to undertake the following assessments:

- (1) To act as a referee to us as guest editors of the special issue of the *Journal of Applied Econometrics* (JAE) and tell us whether each report is of sufficient quality to be published in the JAE. We have a severe space constraint in the Special Issue, which means that we are aiming at an average of 13 journal pages per report. Therefore we would be particularly grateful for your advice as to which parts of each report are particularly interesting and worth publishing.
- (2) To provide detailed comments for each report on how to improve the presentation, etc. The idea here is *not* to impose your ideas on the author on how the problem should be approached. Rather do we hope you will help the authors to clarify their own ideas, to improve the presentation, to suggest possible cuts (given space constraints), etc.
- (3) To act as comparative assessor for the three reports. That is, to analyse how the three reports differ and possibly why. We emphasize that the experiment is not a competition with a winner or a prize. Rather, it is a field trial in applied econometrics, designed to assess both the degree to which econometrics has 'progressed' since Tobin's article, and also the performance of different approaches to the applied problems we set as tasks. In this respect, we would all like to understand why report A gets different results from report B. The assessors will get sufficient time during the workshop to present their assessments and we would like to publish a 2- or 3- page edited version of your comments in the Special Issue.

7. THE WORKSHOP

The workshop took place in Tilburg from Sunday 15 until Tuesday 17 December 1996. Each of the eight teams was represented by at least one author, except Professor G. S. Maddala who

could not participate due to health problems and his co-author due to visa problems. All eight assessors were present, as was, by special invitation, Professor James Tobin. The eight papers and the written assessments were distributed to all participants at the beginning of the workshop. Thirty-three people attended the workshop:

Teams:

CBS	Hans van Driel / Venuta Nadall / Kees Zeelenberg
Texas	Heather Anderson / Farshid Vahid
Leamer	Edward Leamer
Maddala	—
Maastricht	Denis de Crombrughe / Franz Palm / Jean-Pierre Urbain
Finland	Leif Nordberg / Gunnar Rosenqvist
Dundee	Xiaming Liu / Peter Romilly
Tennessee	Peter Bearse / Hamparsum Bozdogan / Alan Schlottmann

Assessors: See Section 6

Others:

Richard Blundell	University College London
Bert Hamminga	Tilburg University
Marco Hoerberichts	Tilburg University
Kevin Hoover	University of California, Davis, USA
Franc Klaassen	Tilburg University
Jan Magnus	Tilburg University
Mary Morgan	University of Amsterdam and London School of Economics
Wiebe Siegert	Tilburg University
James Tobin	Yale University

We started on Sunday after lunch. After a brief welcome by Jan Magnus, three reports were presented (CBS, Texas, and Leamer). Each report was given 70 minutes: 30 minutes for the author(s) to present the report, 20 minutes for the three assessors to give their comments, and 20 minutes for a reply and general discussion. Each presentation was followed by a 20-minute interval. On Monday, following the same format, Maddala and Maastricht presented their results. Since nobody from the Maddala team could be present, Michael Wickens kindly agreed to present the report in their place. Then followed a one-hour general discussion on the experiment as a whole, where the participants gave their views on the strong and (more often) weak points of the set-up. After lunch, the three final reports were presented: Finland, Dundee, and Tennessee. On Tuesday, Kevin Hoover gave his general views about the experiment. This was followed by James Tobin's address, and a 90-minute general discussion chaired by Richard Blundell. Mary Morgan closed the workshop before lunch.

8. AFTER THE WORKSHOP

In December 1996, immediately after the workshop, we set the parameters for reporting the results of the experiment as a whole. This involved the collaboration of us, as organizers of the experiment and guest editors for this Special Issue of the *Journal of Applied Econometrics* (and editors of an enlarged book version), the assessors who assessed the reports for us, and the participating teams.

Our main constraint was space, and our main publication considerations were accurate reporting of the experiment in conjunction with interest, variety and quality of the experimental reports. On the basis of the assessors' comments and general discussion at the workshop, we first decided to drop our policy question task (item 3.4 above). It was generally agreed that the policy task had been poorly conceived by us. We also emphasized (again) that all reports should have a short logbook. On the basis of the assessor's comments on the intrinsic interest of the various approaches to the other experimental tasks and the quality of the reported work, we set individual page/word lengths for the revised reports, and indicated, for some reports, those sections we thought it might be appropriate to cut. We decided to publish six reports in the special issue, and all of them in the enlarged volume.

We wrote to the participating teams: 'The most important point, which we have stressed before, is that this is not a normal academic paper, but a report on an experiment. We ask you therefore to stick to the spirit of our enquiry by keeping to your original reporting both in content of the choices you made and results obtained as well as to the style of reporting. Of course you should correct typos and obvious errors in your reporting. You will need to make some cuts due to the restrictions of length imposed by the Journal. You will also want to take account of the assessors' comments where they suggested clarification of your procedures were needed. But please do not redo the work or rewrite your paper as if the assessors' comments were referees' reports. Recalculated results or changed procedures can be reported in your published "reply" to the assessors. For example, if in your first report you used expenditure rather than income data, or you used nominal income rather than real income — these were your choices and you should leave them intact. You can come back to this point in your "reply" and provide your response.'

The revised reports were received by 1 March 1997. Although most teams adhered to the spirit of the experiment in making revisions, one participant (Leamer) did substantial reworking on his time-series results as these were queried at the workshop. Another team (Dundee), at our suggestion, corrected some calculation mistakes suspected by one of the assessors. Both these changes, and other minor ones, are clearly reported by authors in their final reports. Because of our requests to cut the size of some reports, several teams were obliged to present fewer results and in more succinct form, necessitating a number of new tables.

After a final check of the team reports, we summarized the changes made for every report, and sent all reports back to the assessors on 19 March. We asked them to revise their assessment comments on each report in the style of the discussion published after papers in the *Journal of the Royal Statistical Society* (an ideal role model, since this was the place of publication for Tobin's original article). We also asked our assessors for any comparative comments and remarks on the experiment itself for publication in the enlarged book report of the experiment.

The final round in May 1997 was to invite the participating teams to write brief responses to the assessors' comments again taking the *Journal of the Royal Statistical Society* as the model.

REFERENCES

- Blank, R. M. and A. S. Blinder (1986), 'Macroeconomics, income distribution, and poverty', in S. H. Danziger and D. H. Weinberg (eds), *Fighting Poverty: What Works and What Doesn't*, Harvard University Press, Cambridge.
- Blinder, A. S. and H. Y. Esaki (1978), 'Macroeconomic activity and income distribution in the postwar United States', *Review of Economics and Statistics*, **60**, 604–9.

- Carroll, C. D. and L. H. Summers (1991), 'Consumption growth parallels income growth: some new evidence', in B. D. Bernheim and J. B. Shoven (eds), *National Saving and Economic Performance*, NBER Research Project Report, University of Chicago Press, Chicago.
- Chetty, V. K. (1968), 'Pooling of time series and cross section data', *Econometrica*, **36**, 279–90.
- Cutler, D. M. and L. F. Katz (1991), 'Macroeconomic performance and the disadvantaged', *Brookings Papers on Economic Activity*, **2**, 1–73.
- Izan, H. Y. (1980), 'To pool or not to pool: a reexamination of Tobin's food demand problem', *Journal of Econometrics*, **13**, 391–402.
- Maddala, G. S. (1971) 'The likelihood approach to pooling cross-section and time-series data', *Econometrica*, **39**, 939–53.
- Magnus, J. R. and M. S. Morgan (1995a), 'An experiment in applied econometrics: call for participants', *Journal of Applied Econometrics*, **10**, 213–16.
- Magnus, J. R. and M. S. Morgan (1995b), *The Experiment in Applied Econometrics: Information Pack* (54 pages), CentER for Economic Research, Tilburg University.
- Magnus, J. R. and M. S. Morgan (eds) (1998), *The Experiment in Applied Econometrics*, John Wiley, Chichester and New York, to appear.
- Quah, D. (1994), 'One business cycle and one trend from (many) many disaggregates', *European Economic Review*, **38**, 605–13.
- Tobin, J. (1950), 'A statistical demand function for food in the U.S.A.', *Journal of the Royal Statistical Society, Series A*, **113**, Part II, 113–41.