NAKE

Nieuws

volume 9, number 3

October 1997

Netwerk Algemene en Kwantitatieve Economie

Netherlands Network of Economics

NAKE Secretariat

Ms. José C.M. Dijkzeul

Roetersstraat 11

1018 WB Amsterdam

Phone: 020-525-4199 / Fax: 020-525-4254.

(Office staffed on Monday, Tuesday, Thursday, and Friday.)

E-mail: nake@fee.uva.nl

Home Page: http://www.fee.uva.nl/vak_groep/nake

PREFACE

In this *NAKE Nieuws* you find the best reports on the remaining two lecturers of the June workshop which was held in Groningen. **Bjørn Volkerink** (RUG) reports on 'Economic Growth: Lessons from Economic History' by Nicholas Crafts and **Pieter Jelle van der Sluis** (UVA) writes on 'The Economic Approach to Evaluating Social Programmes' by James Heckman.

We have once again managed to line up three lecturers of very high quality for the forthcoming NAKE Workshop to be held December 8-12 at the University of Nijmegen. Elsewhere in this *NAKE Nieuws* you find details on the courses, registration forms, etcetera. Professor **Peter Phillips** (Cowles Foundation, Yale University) will give the econometrics lectures on the topic of 'Trends and Spurious Regressions.' Professor **Ariel Rubinstein** (Tel Aviv University) gives the microeconomics lectures on 'Modelling Bounded Rationality.' The macroeconomics / methodology lectures will be given by Professor **Kevin Hoover** (University of California, Davis).

The second half of our Utrecht program is almost upon us. On the registration form in the middle of this *NAKE Nieuws* (and on our homepage) you find a list of the courses on offer in blocks III and IV. The selection should be broad enough for everybody to find something to her/his liking. Please use the registration form to enrol.

Ben Heijdra

NAKE WORKSHOP

8 - 12 December 1997

University of Nijmegen

During the week from Monday, December 8th to Friday, December 12th, the Netherlands Network of Economics (NAKE) will organize a Ph.D. workshop. Three distinguished economists will teach intensive courses on microeconomics, macroeconometric methodology, and econometrics. Each course consists of five lectures spread out over five days.

Courses

Peter Phillips, Yale University & Cowles Foundation

'Trends and Spurious Regressions'

Ariel Rubinstein, Tel Aviv University

'Modelling Bounded Rationality'

Kevin Hoover, University of California at Davis

'The Methodology of Empirical Macroeconomics'

Register by filling out the form located in the middle of this *NAKE Nieuws* and returning it to the NAKE secretariat **by 14 NOVEMBER at the latest**.

PROVISIONAL PROGRAMME NAKE WORKSHOP
NIJMEGEN, 8 - 12 DECEMBER 1997

Monday December 8	**Tuesday December 9**
10.30 - 11.30 registration/coffee 11.30 - 13.00 Hoover *13.00 - 14.00 lunch* 14.00 - 15.30 Rubinstein *15.30 - 16.00 coffee/tea* 16.05 - 17.30 Phillips *17.30 - 18.30 reception*	09.00 - 11.00 Rubinstein *11.00 - 11.15 coffee/tea* 11.15 - 13.15 Phillips *13.15 - 14.15 lunch* 14.15 - 16.15 Hoover
Wednesday December 10	**Thursday December 11**
09.00 - 11.00 Phillips *11.00 - 11.15 coffee/tea* 11.15 - 13.15 Hoover *13.15 - 14.15 lunch* 14.15 - 16.15 Rubinstein 16.15 - 18.15 private consultations	09.00 - 11.00 Hoover *11.00 - 11.15 coffee/tea* 11.15 - 13.15 Rubinstein *13.15 - 14.15 lunch* 14.15 - 16.15 Phillips *20.00 workshop dinner*
Friday December 12	
09.00 - 10.30 Phillips *10.30 - 10.45 coffee/tea* 10.45 - 12.15 Rubinstein *12.15 - 13.15 lunch* 13.15 - 14.45 Hoover	

REGISTRATION

Participation in the workshop is free for AIO's/OIO's of the institutions participating in NAKE, and includes tea, coffee, lunches, reception, as well as dinner on Thursday. The participants cover the costs of accommodation, breakfast, and the Course Readers. These costs, together with travel expenses, can however be declared at the faculties. Hotel rooms are available in the **Hotel Atlanta** and the **Apollo Hotel.** It is possible to share a room. Approximate prices are ***f* 95,- to *f* 130,-**.

The number of participants in the workshop is limited. NAKE students have precedence, and the date of receipt of the registration form is also taken into consideration. Since firm arrangements must be made for lunches, dinner, accommodation etc., we would like you to notify the NAKE secretariat in case of any alterations to your plans. You register by filling out the form on the middle page (as completely as possible) and returning it to the NAKE secretariat **by 14 NOVEMBER latest**. Upon registration you will receive written confirmation together with readers for the courses, hotel information, etc.

A number of AIO's/OIO's will be presented with the NAKE diploma during the workshop dinner on Thursday evening.

PRIVATE CONSULTATIONS

During the workshop it is possible for participants to have a one-to-one talk with one (or more) of the lecturers. Students who wish to confer with one of the lecturers about their research are invited to hand in a brief (one-page) description of the research (-proposal) they would like to discuss. Each consultation will be approximately 30 minutes.

METHOD OF ASSESSMENT AND CREDITS

The NAKE workshops are obligatory for all first- and second-year graduate students following the NAKE programme. Hence, each student must attend at least four workshops. For three workshops the student must submit a written summary of the lectures of one course. This report must be based both on the notes taken during the workshop and on the assigned literature. These reports are assessed by the organiser(s) of the workshop. All (NAKE) students are expected to attend all sessions on offer during the workshop.

With regard to study intensity, participation in the workshop (including the assessment by means of the written report) is worth 2 "Study Points" (SP); 1 SP = 40 hours.

ADDRESSES AND INFORMATION

Location:
Zalencompex Valdin
van Peltlaan 4-8
6533 ZM Nijmegen
Phone: 024-355-6902
Fax: 024-354-1990

Information:
NAKE Secretary, Ms. José Dijkzeul
Roetersstraat 11, 1018 WB Amsterdam.
Phone: 020-525-4199,
Fax: 020-525-5280,
E-mail: nake@fee.uva.nl
Home page: http://www.fee.uva.nl/vak_groep/nake

Hotels:

Hotel Atlanta	Hotel Apollo
Grote Markt 38-42	Bisschop Hamerstraat 14
6511 KB Nijmegen	6511 NB Nijmegen
Phone: 024-360-3000	Phone: 024-322-3594
Fax: 024-360-3210	Fax: 024-323-3176

Local organizers:
Robert Paul Berben, 024-361-6172
Jolanda Peeters, 024-361-2783

COURSE OUTLINES DECEMBER WORKSHOP

Trends and Spurious Regressions

Peter C.B. Phillips

(Cowles Foundation, Yale University)

Outline

Spurious regressions and nonsense correlations have a long history in statistics. The literature abounds with humorous examples, and the pitfalls of regression and correlational studies are now common knowledge, even to non-specialists. The phenomenon has attracted serious statistical analysis in recent years, but is still only partially understood. The commonality of trending mechanisms in economic data make econometric regressions especially vulnerable to spurious empirical relationships.

These lectures will develop tools for analysing and understanding protypical spurious regressions between trending time series in econometrics. Asymptotic methods of analysis and approximating model methods will be used to develop the theory and to explain results. Extensions of the theory to panel data situations will be given. Some new interpretations of the phenomena will be put forward and the implications for empirical research will be discussed.

Trends and Trend Elimination

Canjels, N. and M. Watson (1997). 'Estimating deterministic trends in the presence of serially correlated errors,' *Review of Economics and Statistics,* May, 184-200.

Granger, C. W. J. and P. Newbold (1974). 'Spurious regressions in econometrics,' *Journal of Econometrics* 74, 111-120.

Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series.* New York: Wiley.

Phillips, P. C. B. And C. C. Lee (1996). 'Efficiency gains from quasi-differencing under nonstationarity.' In P. M. Robinson and M. Rosenblatt (eds.), *Essays in Memory of E. J. Hannan,* Springer Verlag: New York.

Phillips, P. C. B. and J. Y. Park (1988). 'Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors,' *Journal of the American Economic Association*, 83:401, 111-115.

Background limit theory

Phillips, P. C. B. (1988). 'Multiple regression with integrated processes,' in N. U. Prabhu, (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics,* 80, 79-106.

Phillips, P. C. B. (1989). 'Partially identified econometric models,' *Econometric Theory*, 5, 181-240.

Phillips, P. C. B. and V. Solo (1992). 'Asymptotics for linear processes,' *Annals of Statistics*, 20, 971-1001.

Asymptotics for Spurious Regressions

Phillips, P. C. B. (1986). 'Understanding spurious regressions in econometrics,' *Journal of Econometrics*, 33, 311-340.

Durlauf, S. N. and P. C. P. Phillips (1988). 'Trends versus random walks in time series analysis,' *Econometrica*, 56, 1333-1354.

Mathematical Models of Spurious Regressions

Phillips, P. C. B. (1995). 'Nonstationarity and cointegration: Recent books and themes for the future,' *Journal of Applied Econometrics*, 10, 87-94.

Phillips, P. C. B. (1996). 'Spurious Regression Unmasked,' Cowles Foundation Discussion Paper, Yale University, No. 1135.

Extensions to Panel Data

Phillips P. C. B. and H. Moon (1997) Linear Regression Limit Theory for Nonstationary Panel Data", University of Auckland Discussion Paper.

Modelling Bounded Rationality

Ariel Rubinstein

(Tel Aviv University)

The topic will be 'modeling bounded rationality.' The basic reference is Rubinstein (1997), a book which will be out in December (MIT Press). The plan of the lectures is as follows (references to the relevant chapters of the book and 1-2 key papers are included per lecture):

1. Introduction

Rubinstein (1997), Chapter 1.
Shafir, E. and Tversky, A. (1995), 'Decision Making' in Osherse and Smith, eds. *Invitation to Cognitive Science: Thinking*, 77-109, MIT Press.

2. Issues in Modeling Bounded Rationality in Decision Making

Rubinstein (1997), Chapters 2, 3.
Rubinstein, A. (1988) 'Similarity and decision Making Under Risk', *Journal of Economic Theory* **46**, 145-153.
Geanakoplos, J. (1989), 'Common Knowledge, Bayesean Learning and Market Speculation with Bounded Rationality', *Journal of Economic Perspectives* **6**, 58-82.

3. Imperfect Recall

Rubinstein (1997), Chapter 4
Piccione, M. and Rubinstein, A. (1997) 'On the Interpretation of Decision Problems with Imperfect Recall', *Games and Economic Behaviour* **20**, 3-24.

4. Modeling Bounded rationality in Games

Rubinstein (1997), Chapter 7.
Osborne, M. and Rubinstein, A. (1997), 'Games with Procedurally Rational Players.' *American Economic Review* (forthcoming).

5. Complexity Considerations in Repeated Games
Rubinstein (1997), Chapters 8-9
Rubinstein, A. 'Finite Automata Play a Repeated Prisoner's Dilemma.' *Journal of Economic Theory* **46**, 146-153.

The Methodology of Empirical Macroeconomics

Kevin Hoover

(University of California, Davis)

COURSE DESCRIPTION

Most economic methodologists divide into two camps. On the one hand, traditional methodology is regarded as prescriptive, directing the practice of economists on the basis of philosophical first principles. On the other hand, the recent tendency among methodologists reflects a reaction to the pretensions (arrogance, perhaps) of the methodologist telling the subject-matter specialist how to proceed. The recent approach is more descriptive. It accepts the autonomy of the economic practitioner and tries to understand the internal principles on which he proceeds or to reconstruct the social or rhetorical strategies that inform his practice without giving advice. The recent tendency is often agreeable to economists who wish to dismiss methodology as irrelevant and a waste of time. In this course, we shall explore an alternative, pragmatic approach.

The leading notion of this approach is that a relevant methodology must start with the genuine concerns of a particular discipline. Careful analysis of those concerns, however, often reveals problems of varying degrees of generality that transcend the concerns and the methods of a single discipline. On this view, methodology and the philosophy of science are not disjoint disciplines from economics, but continuous with it, though operating at a higher level of abstraction. Methodology may sometimes help to solve genuine problems in a particular specialty and is, in that sense, prescriptive. To remain grounded and relevant, these lectures begin with the consideration of concrete issues in macroeconomics. How should macroeconomic models be formulated? What is the relationship between theoretical models and empirical tests of those models and other empirical observations? What is the relevance of macroeconomic models to policy? Should macroeconomics be viewed as a special case of microeconomics?

Questions like these raise issues with a long history in the philosophy of science: the scope and nature of economic laws, the role of idealizations, methodological individualism, and the problem of causality. The first lecture in this course, will by way of introduction, consider in detail a particular macroeconomic model and its empirical implementation, highlighting the methodological questions it raises. The second to fourth lectures will take a deeper look at some of those problems. The final lecture will consider some general methodological questions in light of previous four lectures.

LECTURES

1. Some Methodological Problems in Macroeconomics
2. Are There Macroeconomic Laws?
3. Does Macroeconomics Need Microfoundations?
4. Causality in Macroeconomics
5. Pragmatism, Realism and the Practice of Macroeconomics

READINGS BY LECTURE

Background

Blaug, Mark. *The Methodology of Economics: Or How Economists Explain*, 2nd edition. Cambridge: Cambridge University Press, 1992.

Hausman, Daniel. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press, 199? (Appendix)

Hacking, Ian. *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.

Lecture 1

Pissarides, Christopher A. "Loss of Skill During Unemployment and the Persistence of Employment Shocks," *Quarterly Journal of Economics,* November 1992.

Lecture 2

Carl Hempel. "Laws and Their Role in Scientific Explanation," reprinted in Boyd et al., editors. *The Philosophy of Science*. Cambridge, MA: MIT Press, 1991.

Cartwright, Nancy. "Where Do Laws of Nature Come From?" *Dialectica*, 1997.

Kydland, Finn and Edward Prescott. "Time to Build and Aggregate Fluctuations," *Econometrica*, November 1982.

Nowakowa, Izabella and Leszek Nowak. "On Correspondence between Economic Theories," in Bert Hamminga and Neil B. De Marchi, editors. *Idealization VI: Idealization in Economics*, Poznan Studies in the Philosophy of Science, no. 38, 1994.

Cools, Kees, Bert Hamminga, Theo A. F. Kuipers. "Truth Approximation by Concretization in Capital Structure Theory," in Bert Hamminga and Neil B. De Marchi, editors. *Idealization VI: Idealization in Economics*, Poznan Studies in the Philosophy of Science, no. 38, 1994.

Lecture 3

Lucas, Robert E., Jr. "Econometric Policy Evaluation: A Critique," in Karl Brunner and Allan H. Meltzer (eds) *The Phillips Curve and Labor Markets*, vol. 1 of Carnegie--Rochester Conference Series on Public Policy, Amsterdam: North Holland, 1976. Reprinted in Lucas Studies in Business Cycle Theory, Cambridge, Massachusetts: MIT Press, 1981.

Kirman, Alan P. "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives*, Spring 1992.

Janssen, Maarten. *Microfoundations: A Critical Inquiry*. London: Routledge, 1993.

Hartley, James E. *The Representative Agent in Macroeconomics*, London: Routledge, 1997

Mäki, Uskali. "Scientific Realism and Some Peculiarities of Economics," Boston Studies in the Philosophy of Science, 1994.

Lecture 4

Simon, Herbert A. (1953) "Causal Ordering and Identifiability," in: *Models of Man*, ch. 1. New York: Wiley 1957.

Simon, Herbert A. and Nicholas Rescher. (1966) "Cause and Counterfactual," *Philosophy of Science* **33**, 323-340.

Granger, C.W.J. (1980) "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control* **2**(4), November, 329-352.

Cartwright, Nancy. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Lecture 5

Peirce, Charles S. "The Fixation of Belief," in Charles Hartshorne et al., editors. *Collected Papers of Charles Peirce*, vol. 5.

Quine, Willard. "Two Dogmas of Empiricism," in *From a Logical Point of View*, 2nd edition. Cambridge, MA: Harvard University Press, 1961.

Lawson, Tony. "A Realist Theory for Economics," in Roger Backhouse, editor. *New Directions in Economic Methodology*. London: Routledge, 1994.

SELECTED RELEVANT PUBLICATIONS OF KEVIN HOOVER

"Abduction and the New Riddle of Induction," *The Monist*, July 1980 (with J. Harris). Reprinted in Eugene Freeman, editor, The Relevance of Charles Peirce. LaSalle, IL.: Hegeler Institute, 1983.

The New Classical Macroeconomics: A Sceptical Inquiry. Oxford: Basil Blackwell, 1988.

"The Logic of Causal Inference: Econometrics and the Conditional Analysis of Causation," *Economics and Philosophy*, October 1990.

"Scientific Research Program or Tribe? A Joint Appraisal of Lakatos and the New Classical Macroeconomics," in *Appraising Economic Theories: Studies in the Application of the*

Methodology of Research Programs, Mark Blaug and Neil de Marchi, editors. Aldershot: Edward Elgar, 1991.

"The Causal Direction Between Money and Prices: An Alternative Approach," *Journal of Monetary Economics*, June 1991.

"Causation, Spending and Taxes: Sand in the Sandbox or Tax Collector for the Welfare State?," *American Economic Review*, March 1992 (with S. Sheffrin).

"Causality and Temporal Order in Macroeconomics or Why Even Economists Don't Know How to Get Causes from Probabilities," *British Journal for the Philosophy of Science*, December 1993.

"Econometrics as Observation: The Lucas Critique, Causality and the Nature of Econometric Inference," *Journal of Economic Methodology*, June 1994.

"Pragmatism, Pragmaticism, and Economic Method," in *Contemporary Issues in Economic Methodology*, Roger E. Backhouse, editor. London: Routledge, 1994.

"Six Queries on Idealization in an Empirical Context," *Poznan Studies in the Philosophy of Science and the Humanities*, 1994.

"Post Hoc Ergo Propter Hoc Once More: An Evaluation of 'Does Monetary Policy Matter?' in the Spirit of James Tobin," *Journal of Monetary Economics*, August 1994 (with S. Perez).

"Money May Matter, But How Would You Know?" *Journal of Monetary Economics*, August 1994 (with S. Perez).

" Comments on Cartwright and Woodward: Causation, Estimation, and Statistics," in *On the Reliability of Economic Models: Essays in the Philosophy of Economics*, Daniel Little, editor. Dordrecht: Kluwer, 1995.

"Why Does Methodology Matter to Economics? A Review Article," *Economic Journal*, May 1995.

"In Defense of Data Mining: Some Preliminary Thoughts," in *Monetarism and the Methodology of Empirical Economics: Essays in Honor of Thomas Mayer*, Kevin D. Hoover and Steven M. Sheffrin, editors. Aldershot: Edward Elgar, 1995.

"Is Macroeconomics for Real?" *The Monist*, 1995.

"Facts and Artifacts: Calibration and the Empirical Assessment of Real-Business-Cycle Models," *Oxford Economic Papers*, March 1995.

Nicholas Crafts

Economic Growth: Lessons from Economic History

Report by Bjørn Volkerink, University of Groningen

1. Introduction

In a series of lectures professor Crafts showed the increasing arbitrage opportunities between (endogenous) economic growth theory and economic history. Comparative growth performance can be analysed by using three methods of explanation. One way is to use growth accounting, *i.e.* to state the proximate sources of growth. Another way to 'explain' growth is to look at what underlies growth, *e.g.* what institutions and what policies have caused or retarded growth. This analysis is linked to the (recent) endogenous growth theory. There is also a fundamental way to look at comparative growth experiences, the political economy of growth, *e.g.* by looking at the micro-incentives behind policy formation, market interactions, etc.

The rest of the paper is outlined as follows. The next section discusses some recent advances in growth theory. The third section focusses on the industrial revolution. The fourth section deals with catching up, both of the US versus Britain and of (continental) Europe versus the US. The fifth section focusses on the measurement of living standards in the long run. The sixth section shows what happened to living standards during the British industrial revolution. The last section concludes.

2. Economic History and Economic Growth

The traditional growth models, like the Solow-model, have a number of central features. The main constraint is the presence of decreasing returns to capital accumulation (see Barro and Sala-i-Martin, 1995, pp. 16-24). This leads to a long term growth rate that is independent of investment. Total factor productivity (henceforth: TFP) growth (or: technological progress) pins down the rate of growth of income per head, but this growth is exogenous, it is an input, not a result.

There are basically two types of new growth models (Crafts, 1996). One type tries to explain growth by postulating constant returns to capital accumulation. This can be thought of as the presence of large externalities (see *e.g.* the overview on pages 39-41 of Barro and

Sala-i-Martin, 1995). The other type of analysis claims that there are constant returns to innovative activities, new capital goods are invented regularly (see *e.g.* chapter 6 of Barro and Sala-i-Martin, 1995).

This last type of models, that can be labeled, endogenous innovation models, show that long run growth is proportional to endogenous TFP growth. The incentive to innovate critically depends on the degree of appropriability of returns. These models show that there is more innovation if there is a greater market size, if hold-up problems can be avoided or reduced, if R&D labour is cheaper and if there is the right sort of imperfect competition. Models of endogenous innovation can be be necessary to explain growth, but endogenous innovation is not enough.

These models do however imply that growth can be enhanced (or retarded if the wrong policies are chosen) by competition policy, trade policy, company law, reform of industrial relations, subsidies or taxes on innovative activities and education policy. Historical episodes can be compared by using these factors as yardsticks.

The recent episode of slower growth seems to suggest however that the endogenous innovation models are not very relevant anymore. Growth rates have gone down significantly, and the models cannot explain this. Given the current level of investment, these models are able to explain only about 1% of TFP growth per annum. But, there were some special features that enhanced growth during the afterwar period, that are not relevant anymore, like reallocation and reconstruction. Furthermore, there might be some measurement problems in GDP growth estimates (Crafts, 1996).

3. The Industrial Revolution Revisited

Despite all the fuss about it, the British industrial revolution is a period with only a modest level of growth, at least from today's perspective. The growth rate of real GDP per person averaged about 1.5% per year. It was however a period that showed a large structural change. There was a large move of labour out of the agricultural sector, and a great degree of urbanization. TFP growth has been moderate throughout the period however (Crafts, 1996).

The supply side of the British economy at that time can be best described by a few numbers. Investment as a percentage of GDP has never exceeded 10%. The average direct tax rate was very low, averaging about 2%, where total tax revenue as a percentage of GDP was always less then 12%. The levels of primary and secondary school enrolment were very low. So the economy can be best described as one with low growth potential. Growth estimates diverge widely but on average real GDP per capita growth was about 1% from 1760 to 1801 and about 2 to 3% for 1801 to 1830. The main reasons for the slow TFP growth,

relative to the twentieth century, are uneven technological progress, slow incremental improvements and diffusion of well-known inventions and disincentives to innovative activities. In terms of the modern endogenous innovation models of growth, the small market size also mattered. Furthermore, the patent system did not provide much protection, rent-seeking led to adverse selection of talents and the threat of hold-up problems was very real. The potential of Britain was however higher than that of France. The level of GDP in France was higher, but Britain was a country with a higher degree of urbanization, a higher GDP per head, a higher literacy rate and with lower taxes (Crafts, 1996, pp. 22-23). What seems to have mattered most for the industrial revolution to take place in Britain is the capability of doing microinventions, *i.e.* the terms for coping with (radically) new techniques, and the capability of learning (Mokyr, 1993).

As of 1870 however, Britain's relative economic decline started. It was overtaken by the United States and later on by continental Europe. Some possible (and not mutually exclusive) explanations for this experience are the following. Changes in the nature of technological progress. There was more technological progress and the progress was faster. R&D increased in importance. Furthermore, there were some unique American advantages. The US is a large economy, both in terms of the level of population and income and in terms of a large integrated market area, that opened up opportunities for standardized production. Furthermore the US had a large amount of natural resources. The American exports were relatively intensive in natural resources and were somewhat concentrated around R&D intensive sectors. Finally, it may point to failures to adapt to new circumstances by Britain, or to constraints due to its early take-off. Its institutions, like the regulation of the labour market, seemed to retard change. The production structure had a bias towards 'Victorian' sectors, like textiles, that did not show much progress (Crafts, 1996).

Starting from around the beginning of the twentieth century, TFP growth accelerated in the US to a level of about 2% per year. In Britain it was still low at a level of about 0.5%. Some additional reasons, that are related to endogenous growth theory, for the American growth experience are: the level and dispersion of education and the emergence of large scale work plants. The main feature however appeared to be the capability for learning (Nelson and Wright, 1992).

These points can be illustrated by looking at an endogenous growth model by Lucas (1993). As he points out, the key to economic growth is human capital accumulation. This accumulation takes place both on the job and in a formal context. The keys to accumulation are the learning capabilities of society, the strength of learning spillovers and the sectoral composition of labour. Britain lacked in those key aspects, relative to the US.

4. Catching Up

Another way to explain the prevalence of American leadership as of 1870 is by pointing to its relatively dear labour, and not by stressing learning. It is hypothesized that dear labour in production implies more and better machines and greater labour-saving bias in technological progress. The wage gap was about 60 to 70%, primarily because the wide availability of land, the frontier and high productivity agriculture. But, a model due to Brezis *et al.* (1993) shows that leapfrogging can occur between nations that are in a different phase of development, (*i.e.* that the US can take over the position of the UK) if the wage cost in the backward country are lower. This is the reverse of the conditions that prevailed during that period in the US.

Apart from the wide availability of natural resources and the large internal market, the following factor seem to have played a major role as well for the diverging experiences of America and Europe before World War II. One is that the different learning experiences stem from different factor endowments. If one thinks of production as being a combination of some specific quantities of inputs (*i.e.* the isoquant is not a continuous line but a combination of some points), differing relative factor prices can show up as wholly different production techniques. And hence to different technological progress (Broadberry, 1994). Furthermore, international learning spillovers are weak. There are also indirect effects from industrial relations and managerial capitalism. Corporate capitalism, prevalent in the US, allowed for large scale firms. Moreover, the American technology, the frontier at that time, was not cost effective in Europe for given factor prices. So there was little Europe could do against the American leadership. The labour productivity in American industry sectors was generally higher than in Europe, but it varied widely across sectors. The UK lagged even further behind the rest of Europe because of its tariff structure, that effectively limits the size of its market, its lack of some natural resources and because of its limits to learning (Nelson and Wright, 1992 and Crafts, 1996).

After the second world war, there was catch up in the European growth process relative to the US. Evidence seems to suggest that there was TFP convergence between Europe and the US (Crafts, 1996). Technological transfers for example became much easier. A critical feature was the social capability of Europe (Mokyr, 1993). The postwar golden age was characterized by exceptional economic performance in Europe. Average growth rates were generally above 3% and sometimes even around 5% per annum. There were two processes of catch up. The rich European countries generally caught up with the US and the poorer European countries generally caught up with the richer ones. The growth in this episode did however depend on special factors that cannot be repeated, like the reconstruction after the war and some major shifts in economic policy. Moreover, the period was reflected

by the presence of 'social contracts' and trade liberalization. Hold-up problems were much easier to overcome and the effective market size increased (Crafts, 1996).

The recent experience of the NICs in Asia can be compared to the experience of Europe in its Golden Age. In both groups there was TFP and factor convergence. The TFP growth was, however, higher in Europe, whereas the amount of labour and capital inputs increased more rapidly in Asia. In both groups the government played an important role in solving coordination problems. Outward orientation *i.e.* an effective increase in market volume, was also a common feature.

But, by about 1973, the Golden Age in Europe had come to an end. Catch up growth had been exhausted, or at least had become less relevant. Capital productivity had fallen, there were decreasing returns to routine investment. The 'endogenous' long run productivity growth stabilized at a level of 1 to 1.5% per annum. Thus, the endogenous growth story is not the whole story before 1973, growth level were much higher than could be explained form the models. Finally, the postwar settlements started to have adverse implications under the conditions prevailing during the 1970s.

In the UK things were even worse than in the rest of Europe. Not only had growth been rather moderate during the Golden Age, but the adverse effects during the 1970s were even worse than in the rest of Europe. Several explanations can be given for this. First of all, the UK had less scope for catch up than most countries. Secondly, the UK faced the difficulties of winning the war and had to cope with an extremely weak balance of payments. Thirdly, it missed some major opportunities. And fourthly, it made more serious policy errors than the peer group, and did so more often. More specifically these errors consisted of *e.g.* believing in the wrong model of economic growth, a neglect of training policy, a general focus on short term policies, the presence of very high marginal tax rates, and failure to reform industrial relations. This can be illustrated by the union system in the UK, that can be characterized as decentralized craft unionism. This made the chances for hold-ups very real (Crafts, 1996).

All in all the process of catch up has not been an automatic process, it critically depended on the policies pursued and on the impact of institutions. The evidence from the UK can serve as a warning device not to neglect these features.

5. The Measurement of Living Standards in the Long Run

There are three main dimensions in thinking about standards of living in the long run. A proxy that is widely used is real GDP per person. But sometimes this is a very poor proxy. In the long run, reduction in hours of market work and falls in mortality have added

substantially to growth of living standards on top of conventional GDP per person (Engerman, 1994). Furthermore, conventional comparisons of economic performance, based on levels and rates of growth of real GDP per person are particularly misleading when comparing *e.g.* East Asia and Europe.

Numerous alternatives have been developed to improve upon real GDP per capita as a proxy for the standard of living. These include the Human Development Index (HDI), imputations for changes in mortality, measures of economic welfare that include changes in market hours worked, a 'quality of life' index, a Gini-coefficient adjusted GDP per person measure and the use of heights as a proxy for standards of living (Costa and Steckel, 1995). Some of these alternative will be dealt with in the remainder of this section.

The HDI is an indicator that is composed of several extra indicators apart from real GDP per capita. It is a weighted average of income, life expectancy, and schooling (that is a weighted average of the degree of literacy and school enrolment). It is compiled by the UNDP.

The adjustment of growth figures for changes in mortality are made by comparing levels of utility from income and mortality rates. Changes in mortality rates can be thought of as being exogenous from income levels since the correlation between both is low. There are however some other (potential) problems with this approach because one has to value gains in life expectancy and the outcomes are fairly sensitive to the assumptions about double-counting and the discount rate.

Some equivalent problems pop up if one tries to adjust GDP figures for changes in leisure and non-market work. The focal point is on changes in hours worked, and everything is valued at the final year wage rate. It does not allow for technological progress in non-market work however, so it tends to understate differences between countries. The data requirements are rather constraining. Some numbers show that growth rates adjusted for both changes in mortality rates and for changes in leisure raise GDP growth figures by about 1%. Both factors are about equally important. The correction for changes in leisure tend to raise European growth figures and tend to lower Asian growth figures.

The 'quality of life' indicator is based upon an ordinal aggregation of the ranking of countries on the relative scores on six different aspects of development: real GDP per capita, life expectancy at birth, infant mortality rates, schooling, political rights and civil rights. The way to compare countries is by Borda-ranking superiority. If all aspects are weighted the country with the highest score is placed on top. This ranking is less restrictive than applying the Pareto dominance principle, that often leads to deadlocks. In comparing Europe and some Asian countries for example, there are no Pareto dominant cases. By applying the Borda rule however, Europe in generally performs better than Asia.

So, as some figures show, it is wrong to look at for instance mortality alone in trying to get a grip on changes in living standard. Adjusting for changes in leisure are important too. The results are rather indecisive however, they vary widely depending on the weighting scheme. Furthermore future changes have to be made by *e.g.* incorporating environmental variables.

6. Living Standards During the Industrial Revolution

One of the main controversies among economic historians is the question what happened to living standard during the early years of industrialization in Britain. Although real GDP per capita rose, living standards may have gone down for some workers during some subperiod. In this section some of the concepts defined above are used to check for this claim. Furthermore, it is explored what policy failures have been made during this period (Crafts, 1996).

At first some numbers. Real earnings had gone up by about ten percent in 1830 as compared to 1770. By 1860 it had gone up by about 24%, relative to 1830. Real GDP per head shows about the same trend. If one looks at the average height of army recruits, however, there is a decline from 1830 to 1860. Life expectancy steadily rises, but decreases a little after 1830. Infant mortality rates fluctuate but generally go down. Adult literacy and the average years of schooling per person go up slowly. There is no change in the indicator of political rights, whereas the indicator of civil rights shows a sharp improvement after 1830 preceded by an initial worsening. All in all, the HDI, the 'quality of life' indicator, the HDI adjusted for inequality and the GDI, the gender related HDI all show increased performance during the period 1760 to 1850, contrary to the height indicator. In terms of the 'quality of life' indicator, the Borda rule suggests continuousw improvement over time, but in terms of Pareto ranking, this does not hold for 1830 and 1850.

It has been argued by some that heights are a good indicator of health, because heights are affected by nutrition and disease, factors that depend on the level of welfare. On the other hand some objections can be raised against the use of heights as an indicator of wealth as opposed to for example income or mortality. One point is that is not entirely clear what health is influenced by. Furthermore health is sensitive to relative price differences. If food is relatively expensive it might be optimal to reallocate consumption in such a way that your height suffers. But this is an optimal choice. Moreover, migration into urban areas, with worse environmental circumstances, can also have an adverse effect on average height, but this is an optimal choice too. Finally, heights are influenced by expenditures that are in GDP, hence double-counting may occur (Costa and Steckel, 1995).

So, we can conclude the following. All results shown are quite sensitive to the weighting scheme at hand. Heights are a valuable diagnostic, but not a good overall index. Heights and HDI/GDI diverge after 1820.

The key policy failure during the industrial revolution in Britain seems to have been the slow development of public health programmes. Like the provision of sanitation, floating water etc. This underprovision can be seen as a classic example of a market failure. The main problem is that it has to financed out of local revenues. The investment could have been made, both in terms of efficiency and in terms of capabilities. Both the heights and the 'quality of life' indicator provide a warning that is not provided by the real GDP per capita 'indicator'.

7. Concluding Remarks

During the lectures, professor Crafts has dealt with two, somewhat distinct topics. A description and analysis of the growth experience of the last two centuries and measurement and description of standards of living. He has shown that the recent literature on endogenous economic growth can be very useful in 'explaining' the historical performance of countries. Although the main focus has been on Britain, lessons can be learned for the rest of Europe an *e.g.* for the NICs in East-Asia. Furthermore he has shown that just looking at GDP and other 'economic' figures can be very deceptive to really understand what has happened to the standard of living.

References

Barro, Robert J. and Xavier Sala-i-Martin, (1995), *Economic Growth*, McGraw-Hill, Inc., New York.

Brezis, Elise S., Paul R. Krugman and Daniel Tsiddon, (1993), 'Leapfrogging in International Competition: A Theory of Cycles in National Technological Leadership', *American Economic Review*, 83, 5, 1211-1219.

Broadberry, S.N., (1994), 'Technological Leadership and Productivity Leadership in Manufacturing since the Industrial Revolution: Implications for the Convergence Debate', *Economic Journal*, 104, 291-302.

Costa, Dora L. and Richard H. Steckel, (1995), 'Long Term Trends in Health, Welfare, and Economic Growth in the United States', *Historical Paper* #76, NBER Working Paper Series on Historical Factors in Long Run Growth.

Crafts, Nicholas F.R., (1996), 'Endogenous Growth: Lessons for and from Economic History', *CEPR Discussion Paper* #1333.

Engerman, Stanley L., (1994), 'Reflections on ''The Standard of Living Debate'': New Arguments and New Evidence', in: J.A. James and M. Thomas (eds.), *Capitalism in Context*, University of Chicago Press, 50-79.

Lucas, Robert E., (1993), 'Making a Miracle', *Econometrica*, 61, 2, 251-272.

Mokyr, Joel, (1993), 'Editor's Introduction: The New Economic History and the Industrial Revolution', in : Joel Mokyr (ed.), *The British Industrial Revolution: An Economic Perspective*, Boulder, Westview Press, 1-131.

Nelson, Richard R. and Gavin Wright, (1992), 'The Rise and Fall of American Technological Leadership: The Postwar Era in Historical Perspective', *Journal of Economic Literature*, Vol. XXX, 1931-1964.

James J. Heckman

The Economic Approach to Evaluating Social Programs

Report by Pieter Jelle van der Sluis,[1] University of Amsterdam

1. Introduction

Professor James J. Heckman's (1944) research has had an enormous impact on the econometric modeling of labour markets. He obtained a Ph.D. degree in 1971 from Princeton for his thesis *Three essays on household labour supply and the demand for market goods*. Since then he has held appointments at Columbia University, the University of Chicago and Yale University. Currently he is affiliated with the University of Chicago. He is a Fellow of the Econometric Society and of the American Academy of Arts and Sciences. Each year he publishes several articles in major scientific journals. At the Groningen NAKE workshop the topic of his talks was *The economic approach to evaluating social programs*. The problem with the evaluation of social programs is the presence of a *selectivity bias*. Selectivity bias arises when a rule other than simple random sampling is used to sample the underlying population that is the object of interest. In the context of social programs we have to think of people who are selected to receive a training or treatment. The selection is made on the basis of their characteristics and not by random sampling. A good survey of the general problem of selectivity bias can be found in Heckman (1987).

2. Econometric Frameworks for Program Evaluation

The following topics were dealt with in the first two lectures. First of all, the goals of the course were given. The main goal is to construct counterfactuals for the evaluation of social programs. Heckman addresses this issue by 1) Developing simple econometric frameworks (Roy model, Extensions matching, Selection models), 2) Developing precisely the economic parameters of interest, and 3) Looking at alternatives.

Heckman started off explaining the framework of the classical Roy model of self selection and earnings inequality (see Heckman & Honoré (1990)). The Roy model is a model

[1]Address for correspondence: Pieter J. van der Sluis, University of Amsterdam, Department of Actuarial Science and Econometrics, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands, phone +31 20 525 4257, fax +31 20 525 4349, email Sluis@fee.uva.nl.

for a *Robinson Crusoe economy* where there are two potential outcomes (Y_1,Y_2). Here Y_1 refers to the *untreated* and Y_2 to the *treated*. Examples of *treatment* are a job training program or a tuition subsidy for college attendance. The gains for an individual are Y_2-Y_1. The problem is that we only observe if a person goes into state 2.

One way to define the relationship is through the mathematical operation of taking conditional expectations, with $E(Y_2|X)=g_2(X)$ and $E(Y_1|X)=g_1(X)$. Presumably, X includes relevant aspects of the training received by trainees as well as background and local labour market variables. Adopting a linear specification we write $E(Y_2|X)=X\beta_2$ and $E(Y_1|X)=X\beta_1$.

The (log normal) Roy model is set up as follows. Income maximizing agents possess two skills $S_1=s_1$ and $S_2=s_2$ with associated skill prices π_1 and π_2. An agent chooses sector one if his earnings are greater there, i.e. $\pi_1 s_1>\pi_2 s_2$. The proportion of working in sector one, P_1, is the proportion for whom the previous inequality is true. After some manipulations we end up with expressions for the conditional expectations. The *log-normal* Roy model arises when the tuple ($\ln S_1$,$\ln S_2$) is assumed to be normally distributed. Details can be found in Heckman & Honoré (1990). The main message is that the pursuit of comparative advantage reduces inequality in log earnings compared to what it would be if persons were randomly assigned to jobs. The Roy model is based on the assumption that the logs of sector-specific skills are normally distributed in the population. Most conclusions are not robust to this assumption. A general class of nonnormal distributions models can be given for which the main conclusion of the Roy model remains valid. The general class of models assumes that skills can be decomposed into two components: a log-concave random variable and an independent, additive component that can be freely specified. Selection depends on the log concave component but not on the other component. About the empirical content of the Roy model the following conclusions can be drawn: Conditions are given that guarantee identification of his model from knowledge of population earnings densities. Identifiability is a necessary condition for consistent estimation of the Roy model on sample data. In the paper Heckman & Honoré (1990) only the necessary first step is investigated. The development of consistent estimators is left for future research. Under Roy's normality assumptions, it is possible to recover underlying skill distributions from a single cross-section of earnings even though only one skill is observed for any cross-section. Thus prior notions about skill hierarchies, selection from the bottom, and correlations among sector-specific skills can all be subject to rigorous empirical tests. No regressors or conventional exclusions restrictions are required to secure identification. It is not necessary to postulate functional forms connecting means or location parameters to covariates. Even if the sector chosen by agents is unknown, it is possible to identify the parameters of the Roy model up to their labels. (i.e. we know the collection of sectorial means and variances but we do not know which element in the collection

characterizes a particular sector). If the earnings in one sector are not observed, it is still possible to identify many parameters of interest.

These strong identification results vanish in a general nonnormal model. In the general case, any cross-section distribution of wages can be rationalized by a model with independent, positively correlated or negatively correlated skills. Hierarchical models of worker skills and inverse hierarchical models can be constructed that fit the data equally well. A general nonnormal Roy model has no empirical content when applied to a cross-section of wage data. The problem of nonidentification in the nonnormal Roy model is essentially the same as the problem of nonidentification in competing risks models in duration analysis.

Access to data from markets with different relative skill prices facilitates identification. With enough price variation, it is possible to recover underlying skill distributions from aggregate data. It is not necessary to know the sectors chosen by individuals. Skill distributions can be recovered even if the earnings are not observed for persons employed in one sector. Panel data greatly facilitate identification.

It is also demonstrated that if independent regressors are available and the assumption is made that they only affect the location of the log-skill distribution, it is possible to substitute cross-section variation in regressors for multimarket variation in skill prices and recover the underlying skill distributions. It is then demonstrated how access to regressors of variation in skill prices solves the identification problem in the Roy model and gives it empirical content. With sufficient variation in skill prices or regressors, notions about skill hierarchies and association among worker skills have empirical content.

The first half of the second lecture was finished by Heckman with an empirical illustration of the Roy model. Some empirical examples were shown where social experiments were used to evaluate programs. Conclusions are that the Roy model is way too simple to explain some basic phenomena. For example the Ashenfelter's dip which is very common in the data on social programs, is not present in the control group. The remaining part of this lecture was devoted to econometric models of outcome and program participation. We will deal with this part of the lecture in the next section.

3. Evaluation of the Welfare State

In the second half of his second lecture and in his third lecture Heckman talked about cost-benefit criteria, distributional criteria and joint distributions. He mainly talked about Heckman and Smith (1997). There is a diversity of preferences regarding the outcomes of public policies that characterize participants in welfare states. This diversity gives rise to a multiplicity of criteria for evaluating policies. Heckman and Smith (1997) have considered

these criteria and present formal analysis of the information required to evaluate public policies under different criteria. The approximations that are required to go from microeconomic evaluations to conclusions about the general equilibrium outcomes of alternative policies are presented. The conditions under which the conventional econometric analysis of *treatment effects* provide part of the information required to conduct general equilibrium cost-benefit analyses, are presented. It is noted that personal evaluations of policies may not coincide with the evaluations useful in the political arena of the welfare state and methods are presented to reveal private or *subjective* evaluations to supplement and complement the *objective* evaluations.

To implement many of the criteria used to evaluate the welfare state requires information on the joint distribution of outcomes across policies. Traditional cost-benefit analysis avoids this problem by assuming that a background social welfare function automatically solves all of the distributional problems of the welfare state. In this case, which is assumed in much of the micro-econometric evaluation literature, simple per capita measures of economic efficiency based on the change in aggregate output attributable to a policy suffice to evaluate the welfare state. However, even in this case it is noted that estimators widely used in the econometric evaluation literature do not provide the ingredients required for a comprehensive cost-benefit analysis. In an empirical analysis, it is demonstrated that when conventional estimators are modified to account for direct costs and the welfare costs of taxation, they produce very different inferences about program impacts than are produced using standard econometric methods. The conditions are presented under which standard econometric estimators provide reliable answers to well-posed general
equilibrium evaluation questions.

Homogeneity in the response to a policy across persons with the same observed characteristics is the central implicit identifying assumption that underlies most widely-used econometric policy evaluation methods. The assumption of response homogeneity greatly simplifies the evaluation problem. Part of the conflict in the estimates produced from different evaluation criteria arises from heterogeneity in impacts of the same program across persons. Evidence from a major social experiment is presented to show that heterogeneity in response to treatment is an empirically important phenomenon.

An evaluation strategy that properly accounts for individual heterogeneity requires more information than traditional econometric evaluation methods. Heckman demonstrated how information about participant self-selection choices and program participation rules aids in identifying the distributions of outcomes across policies and also provides information on personal valuations of program outcomes. Heckman discussed how social experiments and different types of micro data can be used to identify the criteria considered in this paper and how they can be supplemented with additional behavioural and statistical assumptions to

construct all of the criteria. Unless special individual decision rules characterize program participation, these sources of data do not resolve the fundamental evaluation problem that persons cannot occupy mutually exclusive outcome states at the same time.

Heckman applied some of these methods to data from a major job training program. For adult females, he concludes that the program benefited most participants according to the *objective* evaluation criteria based on gross outcomes, but did not benefit a majority of participants according to self-assessments or the revealed preference behaviour of attriters from the program. The disagreement among the alternative criteria highlights the need for providing information about all of them satisfy the different parties in the welfare state.

4. Matching

In his fourth lecture Heckman first gave a review of the Roy model, and then addressed the issue of *matching* as described in Heckman, Ichimura and Todd (1997). Matching is based on the assumption that conditioning on observables eliminates selective differences between program participants and nonparticipants that are not correctly attributable to the program being evaluated. Let I_0 and I_1 denote the set of indices for nonparticipants and participants respectively. To estimate the treatment effect for each treated person $i \in I_1$, outcome Y_{1i} is compared to an average of the outcomes Y_{0j} for matched persons $j \in I_0$ in the untreated sample. Matches are constructed on the basis of observed characteristics X. Typically, when the observed characteristics of an untreated person are closer to those of the treated person $i \in I_1$, using a specific distance measure, the untreated person gets a higher weight in constructing the match. The estimated gain for each person i in the treated sample is:

$$Y_{1i} - \sum_{j \in I_0} W_{N_0,N_1}(i,j)\, Y_{0j},$$

where $W_{N_0,N_1}(i,j)$ is usually a positive valued weight function defined so that for each $i \in I_1$, $\Sigma_{j \in I_0} W_{N_0,N_1}(i,j)=1$ and N_0 and N_1 are the number of individuals in I_0 and I_1, respectively. The widely used evaluation parameter is the mean effect of treatment of the treated for persons with characteristics X: $M(S)=\mathrm{E}[Y_1-Y_0 \mid D=1, X \in S]$, where $D=1$ denotes program participation. For a particular domain E for X, this parameter is estimated by:

$$\sum_{i=I_1} w_{N_0,N_1}(i)\left[Y_{1i} - \sum_{j\in I_0} W_{N_0,N_1}(i,j)\,Y_{0j}\right]$$

where different values of $w_{N_0,N_1}(i)$ may be used to select different domains E or to account for heteroskedasticity in the treated sample. Different matching methods are based on different weighting functions $w_{N_0,N_1}(i)$ and $W_{N_0,N_1}(i,j)$.

A framework was presented to justify matching methods that allow analysts to exploit exclusion restrictions and assumptions about *additive separability*. After that a sampling theory for *kernel-based* matching methods that allows the matching variables to be generated from parametric or nonparametric estimation methods, was presented It was shown that the matching method based on the propensity score does not necessarily reduce the asymptotic bias or the variance of the estimators of $M(S)$. The advantage of using the propensity score is convenience in modeling. When the method of matching based on the propensity scores method is used, one can estimate treatment effects in two stages. First a model that describes the program participation decision is built. Then a model that describes outcomes is constructed. The sampling theory demonstrates the value of having the conditional distribution of the regressors the same for D=0 and D=1. This point is to be distinguished from the requirement of a common support that is needed to justify the matching estimator. Whether the weighting scheme can be developed to improve the asymptotic variance remains to be investigated.

5. Characterising Selection Bias Using Experimental Data

In his fifth lecture Heckman talked again about matching (Heckman, Ichimura and Todd (1997)) and made from this paper a step to the topic of the characterisation of selection bias using experimental data (Heckman *et al.* (1997)). In this paper a framework is developed for combining experimental and nonexperimental data to test the identifying assumptions that justify three widely-used nonexperimental methods of evaluating social programs based on comparison groups: (1) the method of matching; (2) the classical econometric selection bias model which represents the bias solely as a function of the probability of participation P; and (3) the method of difference-in-differences. The conditional measure of bias is decomposed into three components corresponding to (a) differences in the supports of the regressors between participants and members of the comparison group; (b) differences in the shapes of the distribution of the regressors in the two groups in the region of common support; and (c) selection bias, rigorously defined at common values of the regressors for both groups. The

first two components are eliminated by matching on characteristics that are ‘‘close’’ in the two groups. Only the third component - selection bias - remains.

The methods are applied to unusually rich data from the control group of a random experiment on a prototypical job training program combined with a nonexperimental comparison group of nonparticipants. The decomposition reveals that selection bias, rigorously defined, is generally the smallest of the three components of bias as conventionally measured but it is still a substantial fraction of the experimentally-determined impact of the program we study. In the data, matching reduces but does not eliminate the conventional measure of bias. Matching cannot eliminate a nonzero selection bias, rigorously defined, and in fact the method is based on the assumption that it is zero.

The data are consistent with the index sufficiency assumption that underlies the classical selection bias model. It cannot be implemented semiparametrically because the support of P is limited. To apply the method semiparametrically in future evaluations it is necessary to enlarge the support of P for comparison group members so that it matches the full support of participants, $P \in (0,1)$. The data are also consistent with the identifying assumptions required to justify application of a conditional version of the method of difference-in-differences to the evaluation of job training programs for all but low values of P. The data are consistent with the assumptions that justify matching.

The method of matching and the classical selection bias model share one important feature: under the assumptions that justify each method, selection bias $B(X)$ averages out to zero over certain intervals. Matching is based on the assumption that selection bias is zero for all intervals, however small. The tests reject this assumption, which also underlies the regression method that is advocated in the literature.

Figures in the paper show that the estimated selection bias as a function of P is sizeable, especially in the vicinity of P=0. In that neighbourhood, the shape is broadly consistent with the form of the normal selection bias. However the analysis strongly rejects the application of the normal selection bias model of Heckman (1979). Figures also show that the bias estimated using the classical selection bias model and reveal how far estimates from the inverse Mills’ ratio method are from the nonparametric estimates of bias that are displayed. The conditional difference-in-differences estimator is consistent with the index-sufficient model of selection bias and only requires that bias be the same before and after the data of initial enrolment into the program, or at least be the same in symmetric intervals around the date of initial enrolment in the program.

The cross-section bias detected in the analysis is characterize by a *crossing property*. Sizeable negative bias in some cells or intervals is offset by sizeable positive bias in other cells or intervals. A weighted average across cells can reduce the overall bias substantially.

This is why some form of matching reduces the bias in the sample at hand, although it does not eliminate it.

Heckman also gave a demonstration of the substantial benefits of having access to nonexperimental data that (a) place nonparticipants in the same labour markets as program participants; (b) administer the same questionnaire to both groups; and (c) include information on recent labour force status histories. Recent labour force transitions turn out to be more important predictors of program participation than the recent earnings histories emphasized in the analysis of Ashenfelter (1978). Failure to use comparison groups of persons situated in the same labour markets as participants and administered the same questionnaires contributes substantially to the bias as conventionally measured. These sources of bias are empirically more important than selection bias, rigorously defined. Access to recent labour force histories to predict the probability of program participation considerably improves the performance of nonexperimental methods.

These findings enhance the ability to design future nonexperimental evaluations of training programs. Since the JTPA program that is considered is typical of a variety of training programs in place around the world, the lessons from our study apply more generally. Although further testing with larger samples would be highly desirable, Heckman's analysis suggests that semiparametric sample selection bias methods of the sort proposed by Heckman (1980), Cosslet (1991) and Ahn and Powell (1993) are one potentially promising method for evaluating training programs provided that comparable data are collected on nonparticipants and participants located in the same geographic entities and administered the same questionnaire and provided that the support of the distribution of P for nonparticipants is enlarged. Labour force status history variables, local labour market variables and personal characteristics that determine participation (i.e. Z variables) but are excluded from the outcome equations are valid exclusion restrictions for identifying the semiparametric selection model. The temporal structure of the program makes some of the Z and R variables distinct. Another very promising method is the extension of the method of difference-in-differences. Conditioning on P, the bias function $B_t(P)$ tends to be constant over all t except possibly for low values of P in time periods near the date of random assignment or eligibility determination. It is for this reason that both the index sufficient selection model and our conditional version of the method of difference-in-differences are consistent with each other.

Heckman stresses the importance of collecting information on recent labour force status and of designing nonparticipant samples and of designing nonparticipants samples so that the distributions of P have the same support for both participants and nonparticipants. It is essential to get the full support to identify parameters (1) and (2) for the entire population of participants. Using the argument of Heckman (1990), semiparametric identification in the index-sufficient model requires access to data with subsets for which there is no selection for

the non-participants, i.e. $E[U_0|P,D=0]=0$. A full support design solves the problem of estimating program impacts when they depend on P. Lack of common support -comparing the incomparable- is a major source of selection bias as it is conventionally measured. Heckman's evidence leads to a rigorous reformulation of the definition of selection bias so that it excludes bias arising from gaps in the common support and from the differences in the weights applied to participant and comparison group samples over the region of common support. Using a common support and a common set of weights applied to participant and comparison group samples goes a long way toward improving the performance of any econometric evaluation estimator. Table 16A in Heckman *et al.* (1997) clearly demonstrates this point. Column (1) presents the raw bias ($\hat{B}$) quarter-by-quarter and overall using the means for the control and ENP samples. Column (2) shows how the bias is reduced simply by matching to the nearest neighbour using P. Column(3) shows how the imposition of the common support condition improves the performance of the nearest neighbour matching estimator. Quarter-by-quarter, there is a substantial reduction in bias. However, the greater bias in (2) tends to average out over all quarters. Column (4) presents estimates of the bias that arise from local linear matching. Both procedures impose common support and common weighting and both improve over the raw mean or crude nearest neighbour estimators.

Table 16A here

The same phenomenon appears in table 16B for the difference-in-differences estimator. Simple differencing symmetrically before and after the date of random assignment or eligibility determination eliminates person-specific components of bias. Compare column (1) of that table with column (1) of table 16A. Imposing common support and common density

in column (2) generally reduces the quarter-by-quarter bias although, as is true for matching, the overall bias is not much affected. Using regressors to adjust for the bias reduces it further as shown in column (3). Note in comparing tables 16A and 16B that the overall bias from our conditional difference-in-differences estimator and from the matching estimator are of the same order of magnitude. Table 16C reveals that even though the inverse Mills' ratio as typically applied to badly biased (see the estimated in the first column), weighting by a common density $f(P|D=1)$ greatly improves the performance of the estimator, although the bias from using the weighted estimator is still substantial.

Table 16B here

It is instructive to contrast the biases defined over the common support and with a common weighting with the biases defined in the conventional way (e.g. as in Ashenfelter (1978) or LaLonde (1986)). One conventional measure of bias is the OLS estimator π in the model for controls and comparison group members:

$$Y = g(X) + D\pi + U,$$

where $g(X)$ depends on the specification used. The normal selection bias method introduces the inverse Mills' ratio terms into $g(X)$ in conducting a cross-section analysis. The difference-in-differences method uses Y or regression-adjusted Y differenced symmetrically around the date of random assignment or eligibility determination. Estimates of π reveal the bias in the conventional common coefficient model ($U_0=U_1$), where the program impact is

assumed not to depend on X. This estimate of bias combines the three sources bias distinguished in the Heckman *et al.* (1997) paper plus any bias arising from correlation between U_0 and X. In contrast, estimates of the bias that condition on a common support and impose a common weighting of participant and comparison group data produce an estimate of selection bias as rigorously defined in Heckman *et al.* (1997).

Table 16C here

The estimates of π for the different method are presented in table 17 of Heckman *et al.* (1997). Except for the inverse Mills' ratio, the overall biases (π) from the other commonly-used estimators are of the same order of magnitude. All except the inverse Mills' ratio estimator produce biases that are smaller than the raw mean $\hat{B}$. At the same time, all are large relative to the program impact and exhibit substantial variability across quarters. The different sources of bias tend to cancel each other out. This is especially true of the Barnow, Cain and Goldberger (1980) estimator. (Compare column 3 of table 17 with the final column of table 12 in Heckman *et al.* (1997)).

By decomposing the bias π into its components, we determine whether a small estimated π is due to a fortitious combination of offsetting biases or whether each component of the bias is small. Sources of bias such as the failure of common support and discrepancies in the weights across participants and comparison group members depend on the sampling plan used to collect the data for the comparison group and so are likely to vary across evaluations.

The factors generating self-selection are more likely to be similar across evaluations. The focus in Heckman *et al.* (1997) is on the estimation of the stable components of the conventional measure of bias. Knowledge of these components facilitates generalization of the evidence from any one study to other environments, and is more informative about the sources of bias than the measure B or π traditionally used to summarize bias. Heckman's decomposition demonstrates that in the data they use, selection bias, rigorously defined, is large relative to experimentally-estimated program impacts but is small relatively to the conventional measure of bias.

Heckman's analysis highlights the benefits of randomized trials. While the bias is reduced using nonexperimental methods that impose common support and common weighting, it is not eliminated. Experiments avoid the need to specify precise functional forms of econometric models, or to select regressors to appear in outcome or participation equations. Typically, experimental treatment and control groups are in the same location and are administered the same questionnaires. Experiments solve the problem of common support by balancing the distributions of characteristics between treatments and controls and producing an impact estimate for all P values. If a nonexperimental evaluation method has to be used, semiparametric selection bias models estimated on the data with full support for nonparticipants or conditional difference-in-differences estimators fit outside the period immediately surrounding the period of initial participation in the program appear to be promising methods that deserve much further exploration and testing.

References

Ahn, H. and J. Powell (1993), Semiparametric estimation of censored selection models with a nonparametric selection mechanism, *Journal of Econometrics* **58**, 3-29.

Ashenfelter, O. (1978), Estimating the effect of training programs on earnings, *Review of Economics and Statistics* **60**, 47-57.

Barnow B., G. Cain and A. Goldberger (1980), Issues in the analysis of selectivity bias, in *Evaluation Studies Review Annual, Volume 5*, ed. by E. Stormsdorfer and G. Farkas. San Francisco: Sage, pp. 290-317.

Cosslet, S. (1991). Semiparametric estimation of a regression model with sample selectivity in *Nonparametric and semiparametric methods in econometrics and statistics*, ed. by W.Barnett, J. Powell, and G. Tauchen. Cambridge: Cambridge University Press, pp. 175-198.

Heckman, J.J. (1980), Addendum to sample selection bias as specification error, in *Evaluation Studies Review Annual, Volume 5*, ed. by E. Stormsdorfer and G. Farkas. San Francisco: Sage, pp. 970-995.

Heckman, J.J. (1990), Varieties of selection bias, *American Economic Review* **80**, 313-318.

Heckman, J.J.(1987), Selection bias and self-selection. In *The New Palgrave: A Dictionary of Economics*. Edited by J. Eatwell, M. Milgate and P. Newman. Volume III, London: MacMillan.

Heckman, J.J. and B.E. Honoré (1990), The empirical content of the Roy model, *Econometrica* **58**, 1121-1149.

Heckman, J.J., H. Ichimura J. Smith and P. Todd (1997), Characterizing selection bias using experimental data, Forthcoming *Econometrica*.

Heckman, J.J., H. Ichimura and P. Todd (1997), Matching as an econometric evaluation estimator, Forthcoming *Review of Economic Studies*.

Heckman, J.J. and J. Smith (1997), Evaluating the welfare state. Forthcoming in Econometric Monograph Series, *Frisch Centenary*, ed. by S. Storm, Cambridge: Cambridge University Press.

LaLonde, R. (1986), Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* **76**, 604-620.

REGISTRATION FORM

NAKE Workshop 8 - 12 December 1997, University of Nijmegen

Name .....

Department .....

Institute .....

Address .....

Postal code/Place.....

Phone: ..... E-mail: .....

AIO/OIO/Otherwise.....

(Please circle what is applicable)

Would like to have a private consultation with:

Hoover, Phillips, Rubinstein

Will write a report on the lectures by:

Hoover, Phillips, Rubinstein

Will attend the courses by:

Hoover, Phillips, Rubinstein

Will attend:

reception (Monday) yes / no
workshop dinner (Thursday) yes / no
Special Dietary Requirements? yes / no
(give details if applicable)

Wants a hotel room for: Mon/Tues Tues/Wedn Wedn/Thurs Thurs/Fri

Will share a room: yes / no

with.....

Remarks:

Please return this form by **14 NOVEMBER at the latest** to the NAKE secretary.

Registration Form Utrecht Courses
Blocks I and II, Spring 1998

(please write legibly)

Name: .....

Department: .....

University: .....

Address: .....

.....

Phone no. .....

Email: .....

I will attend the following courses (please mark boxes):

Block III

- ☐ 97.04 *Advanced Labour Economics*. Hartog & Teulings
- ☐ 97.14 *The Economics of the Trade Union*. Booth.
- ☐ 97.48 *History of Economic Ideas*. Morgan.
- ☐ 97.12 *Regional Economics, Agglomeration, and the Global Economy*. Brakman & Van Marrewijk.
- ☐ 97.27 *Coordination Problems*. Goyal & Janssen.
- ☐ 97.9A *Environmental Problems and Policy: A Theoretical Introduction*. Folmer & De Zeeuw.

Block IV

- ☐ 97.51 *Experimental Economics and the Design of Mechanisms*. Schram & Van Winden.
- ☐ 97.24 *The Political Economy of Transition*. Ellman.
- ☐ 97.53 *Social Choice Theory*. Peters & Storcken.
- ☐ 97.9B *Growth and Environment*. Withagen & Smulders.
- ☐ 97.68 *Advanced Industrial Organisation*. Van Cayseele & Furth
- ☐ 97.69 *Optimal Investment Contingency Plans of Firms*. Pfann
- ☐ 97.70 *Applied Policy Analysis*. Van Bergeijk c.s.

Please return this form as soon as possible (preferably **before 19 December**) to the NAKE secretary: José Dijkzeul, Secretariaat NAKE, Roetersstraat 11, 1018 WB Amsterdam.