The NEWSLETTER of THE CLIOMETRICS SOCIETY

December 1986 Vol. 2 No. 2

## United Kingdom Group held their Quantitative Economic History Workshop in September

*by Nick VonTunzelmann*
(University of Sussex)

LONDON-This year's Workshop for British cliometricians was organized by Paul Johnson (LSE) and Roderick Floud (Birkbeck), and held at Birkbeck College with about 37 people attending, including three with North American affiliation.

Michael Huberman, one of the latter (Trent, Ontario), launched proceedings with his paper, 'Inventory and Output Fluctuations in Lancashire Cotton Spinning, 1822-52.' Against the R.C.O. Matthews' view that cotton spinning firms from the 1820s worked their mills at full capacity at all times, with inventories stabilizing in regard to output, Huberman argued that mills in bad times drew up agreements to work short-time (say four days a week), so that output could fluctuate rapidly in the short run. Huberman based this argument on a revision of the yarn output statistics, using a more disaggregated measure than Matthews had utilized. Inventories were a reflection of demand shifts and were seen as destabilizing. Some qualms about the consistency of the disaggregated series were voiced in the discussion, and the need for further information on, for example, the number of firms was expressed. Several speakers doubted the importance of quality signals conveyed through prices for the coarser branches of spinning, and thought the calls for short-time working more exhortation than fact. The need to tighten up the definition of stabilization was affirmed.

John Treble (Hull), 'Sliding Scales and Conciliation Boards: A Quantitative Analysis,' used extensive regression analysis to derive wage and income risk for coalminers in various regions of Great Britain in late Victorian and Edwardian times. His conclusion was that, except perhaps in setting maxima and minima, the conciliation boards operated similarly to the older sliding scales. In this sense they could be evaluated as forerunners of Weitzman-like share systems. The main criticism centered around the alleged inadequacies of the Weitzman approach - some suggested something more along the lines of implicit contracts, and looking at employer as well as employee risk. More formal mean-variance analysis of risk was suggested. Several speakers developed views about the interaction between wage payment systems and the rise of trade unionism. Others commented on the contrast with American experience.

Mark Taylor (formerly of the University of Newcastle-upon-Tyne and now moving to the Bank of England), concluded the session with his paper (jointly

*Samuel H. Williamson, Editor*

with P. C. McMahon), 'Long-run Purchasing Power Parity in the 1920s.' It was unusual that his was the only post-1914 paper this year. Taylor employed the new and highly promising technique of cointegrated time series to compare exchange rate versus price index movements between 1921 and 1925 (monthly data). Thus able to abstract from short-run dynamics, Taylor found that PPP worked well for the major currencies, except for the dollar-sterling rate - the latter divergence was blamed partly on general anticipations of Britain's return to gold at $4.86 from mid-1924. Several of the audience asked why PPP should break down only for $-£: was arbitrage not working? Would the £ not have been affected earlier than mid-1924? Others were worried that capital flows could be relegated to short-run status - this particularly influenced the German mark. As well as some specific queries about the technique, there were some general remarks about the adequacy of necessary conditions here.

The following session coupled two similar papers. James Foreman-Peck (Newcastle-upon-Tyne), in his paper, 'Natural Monopoly and Public Policy in the Nineteenth Century: The Cases of Railways and Telegraphs,' claimed that Britain's laissez-faire approach in the earlier 19th century had led to excessive costs of construction and saddled the country with high railway charges for the rest of the century. Comparison with other similar countries indicated that costs could have been reduced 30% by state ownership. Finally Foreman-Peck developed the "dynamic social savings" arising from induced capital formation and growth, to suggest that for plausible elasticities the conventional estimates of social savings might be nearly doubled on account of such dynamic effects. Some ways of improving these latter assumptions about elasticities were suggested in the comments. Alternative approaches to estimation using the pooled data were discussed, such as the SUR method. Foreman-Peck's important contrast between short and long run cost curves was probed. Several people asked exactly why British costs were higher; if, for instance, land prices were the reason, the conclusion about laissez-faire would not hold.

THE CLIOMETRICS SOCIETY
% Department of Economics
Miami University
Oxford, Ohio 45056 USA



The complementary paper by Robert Millward and Robert Ward (Salford), 'The Comparative Performance of Public and Private Gas Enterprises in the Late Nineteenth Century,' followed up one of Foreman-Peck's earlier Papers (*Economic Journal*, 1985). By estimating translog cost functions for 1898 they showed that municipal enterprises were actually 15% cheaper than private gas companies, but this gap virtually disappeared when standardized for factor prices, population density, etc. Discussion focussed on the extent to which the approach assumed, rather than established, cost minimization. Was output exogenous? which inputs were truly exogenous? were the constraints truly outside the control of management (cf. Lazonick etc.)? There was also some discussion of the data.

On Saturday morning, the first paper was delivered by Francois Bourguignon, and summarized his joint paper with M. Levy-Leboyer, 'An Econometric Model of France during the 19th Century,' as published in the *European Economic Review,* 1984. Both historical inclination and data limitations had led them to formulate a model along Lewis' labour-surplus lines. The model gave good results for each subperiod, though the data had been extensively revised as part of the estimation procedure.

Comparison between early and late century subperiods showed a diminishing effect of agricultural fluctuations, but the mid-century period of income growth was unexpectedly characterized by a high growth of demand for food. Queries were raised about certain components of the model, e.g., over food export prices, foreign demand, and returns on foreign investment. Wider anxieties about the structure of the model were not really voiced, but there was concern about possible inconsistencies between demand-side and supply-side measures, and there was some belief that wages needed to be modelled in other ways than as exogenous (data limitations notwithstanding).

Peter Wardley and Norman Gennell (Durham), 'The Anatomy of Late-Victorian and Edwardian Britain: Characteristics of Economic Structure and Corporate Wealth,' looked again at beliefs about the service sector and showed that compared with manufacturing, transport was highly capital-intensive, and even commercial services began by being more capital-intensive in the mid-century period. They showed that measures of capital-intensity have to allow for the possibility of different sectors facing different factor prices. They also argued against Lazonick-Chandler views about the lack of big business in Britain as compared with the USA - with services included the differences were not dramatic. Some of the participants felt that the Chandler hypothesis was concerned more about the structure of large companies rather than sheer size. Others looked at the adequacy of the Wardley-Gennell measures of productivity growth in services, and whether a good record in services was in the longer run self-defeating.

The final session began with research student Edmund Newell's paper, 'Commercial Policy and the Cooper Ore Trade in the 19th Century' (Nuffield College, Oxford), tracing the vagaries of this commercial policy, especially in the 1840s. Much more was involved than lowering and then abolishing duties, because there was also the ending of duty-free smelting in bond, and of copper's exemption from the Navigation Laws. Nevertheless, Newell rejected the views of some economic historians that the changes had led to the spread of competitive copper industries abroad (e.g., North America, Cuba, Chile). The pattern of prices and output was quite complex. Comments urged a more precise analysis of those complexities, including the identification of demand vs supply shifts and assessment of the slopes of the curves. The effects of the changes in the Navigation Laws could be further assessed in similar vein. There was some concern about the longer-run dynamics of interaction among markets and between freight rates and competitive industrialization abroad.

The balance between homeward and outward freight rates had arisen in Newell's paper, and was a major element in Knick Harley's, 'Coal Exports and British Shipping, 1850-1913' (Western Ontario); a further stage in his long interest in this general area. Harley showed that even though coal exports were a significant component of tonnages exported, the contribution to British shipping has been exaggerated. This was because European destinations predominated, and in this sphere the competition from foreign shipping was strong. Benefits, if any, were conferred on producers and consumers rather than shippers. Several of the questions concerned the time patterns thrown up by Harley's extensive reinvestigation of freight rates, and their determinants. The role of exogenous shocks like the Crimean War and of insurance rates was included with these. The balance between outward and homeward rates was queried in the light of triangular and multilateral trade. It was suggested that Harley could tie in his work here on coal freights with his earlier well-known work on diffusion of steamships (which rested on the geographical pattern of landed coal costs) to develop a dynamic model of growth and technical change in this sector.

Overall, there was some disquiet about the predominance of nineteenth-century studies, and about the absence of some notables on the one hand and up-and-coming graduates on the other. Generally there was, however, agreement that the format was satisfactory and would be used again next year at Newcastle-upon-Tyne (organized by James Foreman-Peck). □

## All UC Group has Fall Meeting

*by Samuel H. Williamson*
(Visiting UC -Berkeley)

SANTA BARBARA, Ca.-The University of California group in economic history held its semi-annual meeting on the Santa Barbara Campus in mid-November. Eleven papers were presented and discussed on a variety of topics. Four papers dealt with the development of the Mexican economy, three by scholars invited from that country. The others covered a broad range of topics including medieval coinage to 20th century agriculture in Iran.

Two papers dealt with the time of the Mexican Revolution. Steven Haber (Columbia) presented evidence that the period was not as revolutionary as one might expect with regard to industry and industrial organization. The warring factions did not destroy the large factories of the country, and after the hostilities ended, ownership was still in the hands of the same "industrial barons" that had owned them before the revolution had started. Enrique Cárdenas (U. de la Americas) then told a fascinating tale of how during six days in November of 1916, the Mexican government gave up on its paper money and drove silver and gold coins *into* circulation.

The other two papers on Mexico moved to an earlier era. Cecilia Rabell (Instituto de Investigaciones Sociales) discussed her work on the tithes records for a Mexican parish from 1673 to 1804. Most of the tithes were paid for in grapes, the major crop of the region. While incomes grew for most in the parish, the production by Indians fell, partly due to the epidemics that periodically swept the area during this time. Ulises Beltrán (Crónica Presidencial) presented an economic analysis of the beginnings of Spanish settlements in the New World, 1521-1640. He concludes that the system of forced labor imposed at the beginning was the optimal way for the Crown to acquire the colony. When factors had greatly reduced the native labor force, there was an incentive to change the system.

Héctor Lindo-Fuentes (UC -Santa Barbara), the conference organizer, asked in his paper the interesting question of why the Central American Federation failed after receiving independence from Spain in 1821. He painted a picture of divided economic interests which could find no benefits to politically unity. In the discussion, many of the participants expressed wonderment at how they stayed together as long as they did--18 years.

Ali Ferdowsi (UC -Berkeley) talked about the NAR system of agriculture in Iran before the land reforms of the 1960's. He contends that the arrangements were much more complicated than the simple feudal model that it has been thought to be. There were extensive irrigation and market forces determining income shares of the various imputs to production.

The conference's attention was then turned to Unions. The first, by Charles Hickson (UC - Los Angeles) , was a theoretic argument that the guild system was not an exercise in monopolization, but an example of the countervailing power hypothesis. The discussion encouraged the author to take the next step and test the theory with examples. The second paper in this group reported on the rapid unionization of the Swedish economy during its industrialization. Its author, Barbara Mikelson (UC -Berkeley) , hypothesized three causes; economic gain, institutional replacement, and the desire to engage in protest.

William McGreevy (World Bank) gave an overview of the Malthusian trap, past and future. While in the past, a falling farm labor force led to lower fertility, it is possible that government policies can be substituted for this effect.

The conference then heard about Chinese women in California during the nineteenth century. Sucheng Chan (UC -Santa Cruz) pointed out that in the mid-19th century there were very few Chinese women in the US and that they were mostly prostitutes; however, this changed fairly rapidly and by the end of the century the majority of these women lived in family settings. Chan was able to trace this change through a time consuming reading of the

*CONTINUED ON PAGE 8*

**Why Economic Historians Should Stop Relying on Statistical Tests of Significance, and Lead Economists and Historians Into the Promised Land**

Donald N. McCloskey
University of Iowa*

I have recently become a nuisance at conferences and in referee reports about statistical significance. The profession deserves an explanation.[1]

I have taken to asking people who use the notion of statistical significance whether they know what they are doing. Quite a few don't.

Step back for a minute and think through what a "significant" coefficient means. It means that the *sampling* problem has been solved, or at any rate solved well enough to satisfy conventional standards. (John W. Tukey has recently given some reasons for doubting the conventional standards: "Sunset Salvo," *The American Statistician*, February 1986, pp. 72-76.) In other words, *the sample is large enough to assure that if you took another sample it would give roughly the same result.* The sampling variance, which is the population's variance divided by the square root of the sample size, has been driven down to some nice, low figure. As John Venn put it in 1888, at a time when our procedures were a mere twinkle in the statistician's eye, the coefficient (or the mean or the difference between two means or the estimated variance or the R-squared or whatever other statistic we are examining) would probably be "permanent." We would probably come up with the same estimate again.

But a permanent coefficient is not necessarily an important coefficient. It could well be that unusually high corn yields in little Iowa would raise the income of the United States a little, and that a proper regression analysis of income on the Iowa corn yields would show this. A large enough sample of years would make the relationship register, and would make it keep on registering in successive samples. (Never mind what "successive samples" of *years* could possibly mean: that's another problem with statistical significance, a philosophical one I'd like to put aside.) The coefficient would be statistically significant. Yet that it registers and would keep on registering does not mean that it is important. "Statistically significant" does not mean "substantively significant."

What matters is oomph. Oomph is what we seek. A variable has oomph when its coefficient is large, its variance high, and its character exogenous. A small coefficient on an endogenous variable that does not move around can be statistically significant, but it is not worth remembering. Oomph is what we mean when we talk about money being "important" for explaining income per person. The Iowa corn yield certainly does affect average national income, but has little oomph because the coefficient is low. Likewise, the existence of oxygen in the atmosphere certainly does affect combustion, but it does not vary enough to give it oomph in an explanation of why the house burned down. The stock of money in the hands of Iowa Citians certainly does determine their expenditures, but because it is entirely endogenous it has no oomph.

Statistical significance, which now guides a large part of the intellectual life of economists, has nothing to do with oomph. It implies, to repeat, that you have acquired some control over sampling error as a source of doubt. Sampling error, though, is seldom the main source

* The author is a Professor of Economics and Professor of History at the University of Iowa.

[1] Cf. *The Rhetoric of Economics* (Madison: University of Wisconsin Press, 1986), Chps. 8 and 9, especially 9; and "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests", *American Economic Review* 75 (May 1985): 201-205.

of doubt. The main source of doubt is whether a variable matters, or whether it matters to such-and-such a degree: what matters is whether foreign prices affected American prices under the gold standard *significantly* (that is, with oomph), or whether American wages affected migration from Europe significantly, or whether social security wealth affected capital accumulation significantly. *Statistical* significance will not reveal this substantive significance, this blessed oomph.

The best way to see the point is to suppose that you really do know what the coefficient is. For sure. God has told you, with no nonsense about confidence intervals; sampling error is zero. The t-statistic is infinite. Well, then: Has the variable got oomph? *You don't yet know*. To find out you have to ask and answer other questions, having nothing to do with statistical significance, such as whether the coefficient is large (how large? Large enough to matter in some conversation of scholars or policy-makers); or whether the variable could vary enough to produce effects you consider important. For most scientific questions the answer that across successive samples that have a nice, random character the coefficient would be permanent (or statistically significant) is only mildly interesting.

"Mildly interesting" is not the same as "not interesting at all." Occasionally an economist will have a genuine sample and because of its small size will have a genuine worry about the sampling problem. But mainly our problems have nothing to do with sampling error. They have to do with other statistical problems (bias, for example: see Leamer's "Let's Take the Con Out of Econometrics") or, most commonly, with oomphelimity.

At this point I need to treat some objections:

[The regressor is confident.] "Statistical significance is an approximate test of what you call 'oomph.'"

Educate me. Tell me how the permanence of an estimate over successive samples tells how important the variable is. To be sure, large coefficients will *ceteris paribus* have larger significance. But why not look directly at the size, and ask directly whether it is large enough to matter? Why be approximate and irrelevant when you can be precise and relevant? Why put the coefficient through an irrelevant transformation? The calculation of statistical significance fools people into thinking they've solved the central intellectual problem, namely, how important a variable is. But the calculation can't do it. It must be done by us: we must decide how large is large. Tables of t tell us how large is large *with respect to the permanence in sampling*. (Yet even they do not tell us where to set the null hypothesis for the test; this again is a question of substance, not of statistics.) The test does not tell us how large is large with respect to the economic argument in question.

[He looks worried.] "But statistical significance provides a good initial hurdle for the variables. They should at least be statistically significant. Those that survive can be tested later for oomph."

No. There's no reason to make a necessary hurdle out of merely desirable quality -- the quality, remember, of appearing to be permanent within such-and-such bounds, at least so far as sampling error is the problem, as it usually is not. Doing so would be like choosing academic colleagues "first" on the basis of their geniality. Geniality is a desirable quality, Lord knows, but not so desirable that it should head a list of lexicographically ordered "priorities." The procedure would make it impossible to hire a brilliant woman with a slightly sub-par amount of geniality. Anyway, for all the talk of "priorities" in public discourse, lexicographical orderings are irrational. The irrationality is greater when the "later testing" for other qualities is not in fact carried out. In actual, middle-brow econometric practice it seldom is. (See the papers cited earlier for some examples drawn from the *American Economic Review*.) Most economists pack up their statistical package and go home as soon as they find "significant" results "consistent with the hypothesis."

[Beads of sweat appear on his forehead.] "But everyone does it. It must have some survival value in producing good economics. And someone who knows more about statistics than I do must have decided that it is a good practice. After all, the econometrics textbooks and the canned programs and all the papers in the journals are filled with it."

The argument here is from authority. Arguments from authority are not always wrong, though this one seems to be. I do not know why economists and quantitative historians have misread their statistics books. It would make a good paper on the rhetoric of econometrics to trace the literature back to the authoritative turnings. (I record the impression that the reliance on significance in dropping variables is not usually recommended in so many words by textbooks, but in practice has figured more heavily as computers have become cheaper.) One can merely quote authorities in reply, and note that the authorities are of the best sort. Again I refer to the articles mentioned above and the works cited there: for instance, the article by William H. Kruskal (past president of the ASA, etc., etc.), "Statistical Significance" in the *International Encyclopedia of Statistics* (1978; and an earlier version in the *International Encyclopedia of the Social Sciences* (1968)); or the elementary book, *Statistics* (pp. 501, A-23, and *passim*), by David Freedman, Robert Pisani, and Roger Purves (well-known statisticians, youngish turks, etc., etc.). The point has been well known since the early days of modern statistics. Only 3.14159% of economists seem to be aware of it (a short list would include some specialist econometricians such a Griliches and Leamer and a few amateurs such as Arrow and Mayer; I learned it from Eric Gustafson).

[He loosens his tie, sweat dripping from his nose.] "But there's nothing else to do. I want to use statistical procedures. What do you propose to substitute? How will I fill up my days?"

Fill them up with statistical calculations that are to the point. Find out what people consider to be a large coefficient and then see if your data show it. Do sensitivity analysis. Bend over backwards to see how robust your argument is. Encompass your opponent's model with yours, showing how his results follow as special cases of yours. Take collecting "data" seriously (the word means "givens" in Latin: we should prefer "capta," things *taken*). There's plenty of useful econometric work to be done (see Sims, Leamer, Hendry, *et al* among econometricians, and Mosteller, Tukey, Hogg, *et al* among statisticians) that does not rely on the misuse of statistical significance.

[He is quaking nervously and his palms are wet. But at once he stops and cools. A smirk spreads over his face. He has found peace.] "To hell with you. As long as editors keep publishing articles that misuse statistical significance I'm going to keep on submitting them. I've got a career to run."

Shame on you. The argument is immoral. Our custom of forbidding talk about morality is strong among economists, some of whom think that the model of selfish behavior is in fact a set of suggestions about how to behave. But there's no two ways about it: it's immoral to lie, and for a scholar it's a mortal sin. That 96.8584% of editors fall into the group of economists who do not know the difference between statistical and substantive significance does not justify someone who does know the difference in going on pretending she does not.

Scholarship that depends on convenient lies will not last. To put it sharply, it is gradually becoming plain that the econometric work of the past quarter century relying inappropriately on statistical significance (which is most of it, unhappily) has to be done over again.

Economic historians are well placed to do better. We capture our own data, and therefore know that errors in variables are no joke. The intellectual traditions of cliometrics favor self-doubt, which in turn favors a sort of counting call "robust." Above all, we are trying to answer substantive historical questions about particular events, not trying to "test" more or less vague hypotheses above Economic Behavior. □

**All UC Group** *(continued from page 4)*

census manuscripts from 1860 to 1910.

Carlo Cipolla (UC-Berkeley) explained the different ways coins were devalued during medieval times. When the desire was to raise money, the mint would lower the amount of silver in the coin. When the desire was to increase the money supply, it would raise the price of silver. Professor Cipolla enlightened us by passing around several coins from the period under discussion.

Kathleen Biddick (Visiting UC -Los Angeles/Riverside) described her study of Peterborough Abbey and why an English feudal estate would not risk full particpation in the market economy.

As with the British Conference of two months earlier, the conference ended with Knick Harley (Visiting UC-Davis) and his paper on coal exports. He explained that though Britain did ship alot of coal in the late 19th century, this was not the source of a comparative advantage that can be used to explain why Britannia ruled the waves. In many cases the rates collected for coal on the "return trips" covered the cost of loading and unloading, and there was little advantage to hauling coal compared to using sea water for ballast. □

*Abstracts of the Papers are in Section 2*

## Letter To The Editor

*With this Newsletter we initiate a Letters to the Editor section. Everyone should feel free to contribute. The following letter was received from Scott Eddie (University of Toronto) who is spending the year in Vienna.*

I appreciate the Newsletter very much, and was fortunate to get a copy in Bern. Thanks for what was very informative, if at times frustrating, reading.

"Clio appears to be fat, happy, and sassy," write Davis and Engerman in the April Newsletter. And smug, it has been added, by some of the European economic historians with whom I met in Bern at the quadrennial congress. While quite comprehensive as far as it went, the Davis and Engerman piece suffers from the same ethno-cultural myopia with which we Americans and Englishmen have so long frustrated our European compatriots. While D & E do seem to recognize that some work has been done on non-English-speaking countries, their view is limited in this regard to one sentence mentioning France and the Netherlands. What of Germany, the Scandinavian countries, Austria, and Italy? Indeed, what of the whole continent outside of those countries which have regular daily ferry service to England? What of the rest of the world?

There is so much more that they could have mentioned. The may not have known about Roman Sandgruber's pioneering work on the consumer society in Austria, since that was published in German, but what of Good, Komlos, Rudolph, and Gross? What about Rolf Dumke's work on the German Customs Union and German income distribution? Or Steven Webb or Bob Moeller on German agriculture and tariff policy? What about Jon Cohen's work on Fascist monetary and agricultural policy in Italy, or John Munro's on money in the Low Countries in the Middle Ages? Then there's George Peteri on Swedish banking, and Lars Sandberg on several topics. What of the Barkai, Drummond, Gregory & Sailors', *et al* debate on Russian monetary policy and the balance of payments? Or Gregory's book on Russian national income? Or Sims, Wilbur, *et al* on the validity of the Gerschenkron hypothesis about post-Emancipation agriculture? I could go on, but I think the point is made: D & E could haved cast their net far wider, and landed a more varied catch, if they had not so obviously suffered from the American disease. □

# Classifieds

The deadline for submissions to the Classified for the April Newsletter is March 15. Note that this is a place to make free announcements, advertisements or want ads.

The 1987 Cliometrics Conference will be held at the University of Illinois on May 15th to 17th. The deadline for submiting a request to attend or give a paper is February 1st. Those requests should be sent to Cliometrics Conference Secretary, 328 David Kinley Hall, University of Illinois, 1407 W. Gregory Dr., Urbana, IL 61801

Nick von Tunzelmann is now putting out a twice a year listing entitled "Economics Publications in Economic History." This is a listing that includes both publications and working papers in the field that have been published in the previous six month period. If you are interested in receiving a copy send Nick $2.00 or equivalent to cover his costs. His address is Science Policy Research Unit, University of Sussex, Mantell Building, Falmer, Brighton, East Sussex, UK BN1 9RF

The 1987 ESRC Quantitative Economic History meeting takes place at Newcastle-upon-Tyne on 18-19 September. Requests to attend and/or to give a paper should be sent to: James Foreman-Peck, Dept. of Economics, The University of Newcastle-upon-Tyne, Newcastle, UK NE1 7RU □