TWO NEW ZEALAND PIONEER ECONOMETRICIANS

BY

PETER C. B. PHILLIPS

COWLES FOUNDATION PAPER NO. 1308

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2010

http://cowles.econ.yale.edu/

New Zealand Economic Papers
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t901392665

Two New Zealand pioneer econometricians

Peter C. B. Phillips [abcd]
[a] Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut, USA [b] University of Auckland, [c] Singapore Management University, [d] University of Southampton,

Online publication date: 24 March 2010

To cite this Article Phillips, Peter C. B.(2010) 'Two New Zealand pioneer econometricians', New Zealand Economic Papers, 44: 1, 1 — 26
To link to this Article: DOI: 10.1080/00779951003614057
URL: http://dx.doi.org/10.1080/00779951003614057

New Zealand Economic Papers
Vol. 44, No. 1, April 2010, 1–26

RESEARCH ARTICLE

Two New Zealand pioneer econometricians

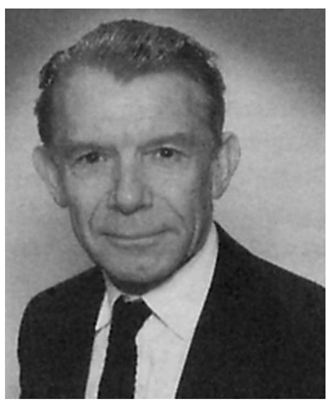
Peter C.B. Phillips*

Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, Connecticut 06520-8281, USA; University of Auckland; Singapore Management University; and University of Southampton

Two distinguished New Zealanders pioneered some of the foundations of modern econometrics. Alec Aitken, one of the most famous and well-documented mental arithmeticians of all time, contributed the matrix formulation and projection geometry of linear regression, generalized least squares (GLS) estimation, algorithms for Hodrick Prescott (HP) style data smoothing (six decades before their use in economics), and statistical estimation theory leading to the Cramér Rao bound. Rex Bergstrom constructed and estimated by limited information maximum likelihood (LIML) the largest empirical structural model in the early 1950s, opened up the field of exact distribution theory, developed cyclical growth models in economic theory, and spent nearly 40 years of his life developing the theory of continuous time econometric modeling and its empirical application. We provide an overview of their lives, discuss some of their accomplishments, and develop some new econometric theory that connects with their foundational work.

Keywords: Aitken; Cramér Rao bound; HP filter; minimum variance unbiased estimation; projection; GLS; Bergstrom; continuous time; exact distribution; LIML; UK economy; pioneers of econometrics

Alexander Craig Aitken



Albert Rex Bergstrom

*Email: peter.phillips@yale.edu

This is a solicited article based on the keynote speech delivered by the author at the New Zealand Association of Economists Conference held at Wellington, New Zealand in June 2009.

ISSN 0077-9954 print/ISSN 1943-4863 online

DOI: 10.1080/00779951003614057
http://www.informaworld.com

Overview

This article tells the story of two New Zealanders who laid some of the foundation stones of the modern discipline of econometrics: Alexander Craig Aitken (1895–1967) and Albert Rex Bergstrom (1925–2005). Sadly, Alec Aitken is little known in the modern economics fraternity, although he is a celebrated figure in New Zealand mathematics circles. Those who studied econometrics prior to 1980 will know of him by association through the once commonly used eponyms 'Aitken's method' and 'Aitken's generalized least squares (GLS)' applied to the estimation procedure that he developed in conjunction with the matrix treatment of linear regression (Aitken, 1934) that is now universal in econometrics.[1] Rex Bergstrom's work is widely known in the international econometrics theatre and he is well remembered in New Zealand for pioneering serious empirical econometric research on the New Zealand economy and for his longstanding contributions as an educator, training successive waves of New Zealand economists over the 1950s and 1960s.

Both Aitken and Bergstrom were born in the South Island of New Zealand, Aitken in Dunedin and Bergstrom in Christchurch. Both worked in New Zealand but ended up spending much of their lives in the UK. Both were athletes in their youth – Aitken winning the high jump and pole vault championships in Otago county as a young man, and Bergstrom winning a bantomweight boxing championship at high school. Both had a lifelong passion for mathematics. Aitken's passion was tempered by wide-ranging artistic interests in music, classical scholarship, literature and poetry. Bergstrom's love of mathematics was tempered by the challenges of economic theory, econometric methodology and empirical modeling that came to occupy so much of his life. Both wrote with elegant simplicity and precision, providing paradigms of economical writing that might be used as exemplars in a work such as McCloskey (2000). Both carried themselves with humility and dignity, emanating distinct academic gravitas: Bergstrom, with the steely look of determination that one might expect in a problem solver, dressed in a charcoal Saville row suit that became his uniform throughout life; Aitken with a look of gentle compassion and faraway contemplation reflecting a devotee of the arts and the abstract preoccupations of a mathematician extraordinaire.

Aitken was a first generation contributor to econometrics. The foundation stones that he laid to the discipline were all completed prior to the methodological work of the Cowles Commission in the 1940s that appeared in the famous Cowles Commission Monograph 10 that was edited by Koopmans (1950). His work on the matrix algebra of regression was contemporaneous with Koopman's (1937) early study of regression statistics. Bergstrom was a second generation econometrician, learning much of his econometrics from Koopmans (1950) and his probability and statistics from Cramér's (1946) famous treatise on the mathematical methods of statistics, which remained one of his lifetime-favorite books. At the time Aitken died, in 1967, Bergstrom's career was in full flight, a new generation of econometrician was emerging, and the discipline had matured to the stage where it was about to give birth to its own scholarly journals. When Bergstrom died in 2005, the subject of econometrics had exploded into numerous sub-disciplines, its tentacles reached into every corner of applied economics, and its methodologies were being used in other social sciences and in the natural sciences as far afield as environmental science and paleobiology.

Alec Aitken

Life and career

Aitken was born in Dunedin to a Scottish father and English mother. His extraordinary mental capabilities were apparent at an early age. There are many fascinating anecdotes. Some of these have become amplified in the retelling. An authoritative source is Kidson's (1968) biographical notes, which are based on personal discussions with Aitken and his family. One example describes his primary school entrance interview. When asked by the headmistress to spell the word 'cat', he ignored the question and proceeded to spell several polysyllabic words, concluding with 'Invercargill'. At Otago Boys High School, where he was dux, he memorized the entire 12 books of the Aeneid. Later in life he became a living legend through astonishing feats of memory and mental arithmetic. Much has already been written of his prowess as an arithmetician – see, for example, the biographies by Kidson (1968) and Garry Tee (1981), and the obituaries by Whittaker and Bartlett (1968) and Silverstone (1968). He is regarded by many as being one of the greatest arithmeticians of the 20th century for whom reliable records exist.[2] In a remarkable BBC radio interview in 1954, he multiplied nine-digit numbers in less than 30 seconds, gave the full repeating decimal representation of the prime number fraction 23/47, and knowing π to a 1000 digits, was able to recite the digits from an arbitrary starting point. As the transcript of the BBC interview (available at www.mental-calculation.com/misc/bbc1954.html) describes it

> Bowden gives Aitken 23/47. Aitken repeats the problem, then reels off the digits '0.48936170212765959446808510638' at a rate of 3 per second. Bowden feels it necessary at this time to reassure the audience that neither calculator has been told the numbers in advance. Aitken continues '63829787234042553191 and that completes the 46 digit period'. Astounded by this, Burt asks Aitken how he knows the digits repeat after 46 digits. Aitken reveals that he arrived at '3191489' and 'remembered 489 started off the expansion'.

In an article on the art of mental calculation, Aitken (1954) describes some of the methods he used in developing these remarkable computing and memory skills. When asked to multiply 987,654,321 by 123,456,789 he explains in a telling remark that is often quoted (e.g. Silverstone, 1968):

> I saw in a flash that 987,654,321 by 81 equals 80,000,000,001; and so I multiplied 123,456,789 by this, a simple matter, and divided the answer by 81. Answer: 121,932,631,112,635,269. The whole thing could hardly have taken more than half a minute.

Topping the national university entrance scholarship exams in 1912, apparently by a wide margin, Aitken entered Otago University in 1913. He gained senior scholarships in Latin, Mathematics, and Applied Mathematics, and went on to take first class honours in Latin and French, and a second in mathematics. The mathematics result was a surprise and a great disappointment. It appears to have been due to a demanding externally set examination that departed egregiously from course content. There was no mathematics professor at Otago and Aitken's only guidance came from tutoring by the mathematics master at his former high school, Otago Boys. To console himself, Aitken took a brief sojourn camping on beaches in Stewart Island and resolved to forgo a career in mathematics.

Accordingly, he assumed the post of a classics master at his old high school over 1920–1923. But his mathematical abilities did not languish. He soon came to the attention of R.J.T. Bell, the new professor of mathematics at Otago, who hired him as a part-time assistant and encouraged him to go on to do doctoral research with Edmund Whittaker, the famous British mathematician at the University of Edinburgh. This move opened up a brilliant mathematical career and a new arena of personal life for Aitken.

Aitken's period at the University of Otago was interrupted by war service. He enlisted in 1915 in the expeditionary army force destined for service in Gallipoli. He served in Gallipoli, Egypt and France, was injured in the battle of the Somme in 1916, and hospitalized for 3 months in the UK before returning to New Zealand.

Aitken was haunted by memories of the war for the rest of his life. Later in life, he crafted his wartime diaries, first written in 1917 while recuperating from his war injuries, into an autobiographical account published in his book (1963) *From Gallipoli to the Somme: Recollections of a New Zealand infantryman*, which was hailed as one of the most compelling and authentic accounts of life in the trenches during the First World War. Integrity shines from its pages in elegant and simple prose that describes the incompetence of combat command, the futile missions, the daily ennui, the abominable living conditions, and the tragic waste of life, capturing a stark reality that defies the glorifications of war so common in fiction and political propoganda. Some accounts in the book have become classic. One episode describes what it is like to line up an enemy soldier in the sights of a rifle, squeeze the trigger, watch the soldier fall, and then contemplate the reality of what has been done to another human being. The book was sufficient to earn Aitken immediate election to a fellowship of the prestigious Royal Society of Literature in 1964.

Aitken began his career at Edinburgh as a doctoral student and remained there for the rest of his life. His doctoral research was conducted under the supervision of Edmund Whittaker, one of Britain's most eminent mathematicians, whose *Course in modern analysis* with (his former student) G.N. Watson was legendary in its own time, is still reprinted as a mathematical classic, and is used in diverse areas of applied mathematics where special functions are employed, including econometrics. Aitken's thesis dealt with the statistical problem of 'graduating' data that is subject to error, which now falls under the general head of statistical smoothing techniques, but at the time was most useful for actuarial calculations. Aitken addressed both the theory and the practical aspects of implementing these smoothing algorithms. His dissertation was considered so distinguished that he was awarded the degree of DSc rather than PhD in 1926. Even before this, he was elected to a fellowship of the Royal Society of Edinburgh at the age of 30 in 1925.

In 1925, Aitken was appointed to a lectureship in Actuarial Mathematics at the University of Edinburgh and so began his academic careeer. Over succeeding years, he became lecturer in Statistics, in Mathematical Economics, and in 1936 was appointed as Reader in Statistics. It would be interesting to explore the archives at the University of Edinburgh to see if there is any detail about the content of his early courses in mathematical economics.

Whittaker retired in 1948 and Aitken was elected to his chair in Pure Mathematics, in which post he remained until his own retirement in 1965. He died in 1967. Kidson (1968) movingly describes two visits in the 1950s to Aitken's home, the former residence of John Ballantyne, Sir Walter Scott's publisher. Aitken entertained Kidson and his wife with a demonstration of his mental arithmetic, some of his own piano and violin compositions, as well as a javelin throw, all undertaken 'with a complete and natural simplicity'.

Aitken received many accolades during his career, including election as a Fellow of the Royal Society in 1936. In 1953 the Royal Society of Edinburgh gave him their highest award, the Gunning Victoria Jubilee prize. He received honorary degrees from the University of Glasgow and the University of New Zealand. The New Zealand Mathematics Society awards the 'Aitken Prize' annually to a student in his honour. A conference in his honour was held at Otago University in 1995 to commemorate the centenary of his birth.

Research and writings

Aitken's main research contributions were in algebra, statistics, and numerical mathematics. Those works that connect most closely with econometrics are discussed in the next section. We briefly mention some other notable contributions here.

As founding joint Editor (with D.E. Rutherford) of the famous Oliver and Boyd series of mathemetical texts, Aitken (1939a, 1939b) personally contributed the first two volumes to this series: *Determinants and matrices* and *Statistical mathematics*. These books were still in common use as texts in the 1960s. The volume on determinants and matrices was reprinted again in 1983. It is as remarkable for its brilliant exposition as it is for its astonishingly broad coverage, taking readers from the novice level through to research topics in a wonderful set of problems, all in the matter of 144 pages. Even now there is no book on matrix albegra to match it. Another famous book (*The Theory of canonical matrices, 1932*) by Aitken (co-authored with H. W. Turnbull) dealt with canonical forms of matrices and is still a classic reference on this subject.

Early in his career Aitken (1925) developed, as part of a more general work that extended Bernoulli's formula for the greatest root of a polynomial equation, a powerful (nonlinear) method of accelerating the convergence of a sequence. The method is still in common use and is known as Aitken acceleration or Aitken's δ^2 method. Interestingly, it is now known that the method was independently discovered much earlier by the Japanese mathematician Seki Kōwa (1642–1708) who used it to calculate π correctly to the 10th decimal place.

In another of his many contributions to numerical mathematics, Aitken (1931) developed a simple procedure for linear interpolation, based on the famous Newton–Lagrange formula, which became known as the Neville–Aitken method. The advantage of this procedure is that it is recursive, enabling users to update the calculation of the interpolating polynomial as successive points are added to the interpolation.

The Royal Society biographical memoir by Whittaker and Bartlett (1968) provides a full bibliography of Aitken's published work and summarizes his many contributions to mathematics and statistics. The archives of the Royal Society record the following citation upon Aitken's election to a Fellowship in 1936:

> Distinguished for his researches in mathematics. Author of 43 papers, chiefly in algebra and statistics. In particular, discovered (1) a theory of duality which links determinantal theory with the combinatorial partition-theory and with group-character theory, (2) the transformation of the rational canonical form of any matrix into classical canonical form, (3) the concept of minimal vectors associated with a singular pencil of matrices, furnishing a method of great value, (4) a theorem of which most of the fundamental

> expansions of function-theory and interpolation theory are special cases, (5) a final solution, theoretical and practical, of the statistical problem of polynomial representations.

Outside his academic work, Aitken was known to his friends and colleagues as a brilliant creative artist and musician. Following his retirement, a special Minute of the University of Edinburgh was recorded for the Senate Archives in 1966, lauding his many accomplishments and describing how . . .

> Edinburgh has always had a great attraction for him and he has resisted the many offers tempting him to go elsewhere. It has most of what he wanted, a congenial job, hills to walk in, and above all, music, concerts, and musical friends. He is a fine (largely self taught) violinist and viola-player, and a very knowledgeable musician . . . Indeed, he is reliably reported as having said that he spent three-quarters of his time thinking about music. This remark reveals how efficiently he must have used the other quarter! He is moreover a creative artist, though only a few intimate friends have been privileged to know his compositions, occasional poems, and the eloquent calligraphy of his manuscript copies of Bach.

Later in life, Aitken (1962) published an impassioned pamphlet against decimalization, proposing in its place the duodecimal system. To a lay audience he described the merits of base 12 arithmetic, explained the duodecimal system with great clarity, and drew analogies to the number systems built around 12, 60, and 240 that have served humanity so well in measurements that range from the time of day to months of the year, and building construction through to currency systems. With characteristic lucidity, Aitken put the position thus:

> With all this, pounds and pence have an advantage which the franc and centime, dollar and cent, metre and centimeter cannot possibly claim, namely the exceptional divisibility of the number 240. This in fact is one of those integers which mathematicians, in that special field called the 'theory of numbers', are accustomed to call 'abundant' . . . Compared with 120 and 144, even with 60, the number 100 is relatively poverty-stricken in this respect – which indeed is why the metric system is a notably inferior one; it cannot even express exactly for example the division of the unit, of currency, metrical or whatever, by so simple, ubiquitous and constantly useful a number as three.

Contributions to econometrics

The part of Aitken's work that contributed directly to econometrics falls into three categories: (i) data smoothing; (ii) matrix mechanics of least squares regression and the associated projection geometry; and (iii) optimality in statistical estimation. These contributions will be discussed briefly below. In addition, according to David (1995), he is the original source for the following commonly used terminology in statistics and econometrics: *minimum variance unbiased estimator* (Aitken & Silverstone, 1942) and *probability function* (Aitken, 1939b).

(i) Data smoothing

Aitken's doctoral dissertation dealt with the graduation of data, following earlier research by his advisor in Whittaker (1923)[3] and Whittaker and Robinson (1924) and work by Henderson (1924). The topic was of major importance at the time, primarily for actuarial calculations, but has remained important in applications ever since. The modern terminology for this field is *data smoothing*, which interestingly is close to Aitken's (1925) dissertation title, as cited in the Mathematical Genealogy Project of the American Mathematical Society (http://www.genealogy.ams.org/). The term was also used in Whittaker (1923).

In economics, these methods have become heavily utilized in the empirical analysis of business cycles. Here the smoother is a numerical trend extraction device (the so-called 'flexible ruler') and the residual is studied for manifestations of cyclical behavior, leading to a trend plus cycle decomposition of a time series. The method is often used to compute new series such as potential GDP and the output gap for implementation in macroeconomic modeling and monetary policy research.

Given data $X^n = \{X_t\}_{t=1}^n$, the Whittaker smoother is the solution to the following numerical problem

$$\hat{f}_t = \arg\min_{ft} \left\{ \sum_{t=1}^{n} (X_t - f_t)^2 + \lambda \sum_{t=m}^{n} (\Delta^m f_t)^2 \right\}, \tag{1}$$

where the primary term $F_n = \sum_{t=1}^n (X_t - f_t)^2$ controls the fidelity (or fit) to the data and the secondary term $S_{mn} = \sum_{t=m}^n (\Delta^m f_t)^2$ imposes a smoothness penalty measured in terms of the magnitude of the higher order differences $\Delta^m f_t$ for some integer $m \geq 2$ where Δ is the difference operator. The idea was simply to get the best fit to the data while smoothing out the irregularities manifest in the differences, so that $\Delta^m f_t$ should be small. As Whittaker (1923) described it,

> ... the problem belongs essentially to the mathematical theory of probability; we have the given observations and they would constitute the 'most probable' values.... were it not that we have a priori grounds for believing that the true values ... form a smooth series, the irregularities being due to accidental causes which it is desirable to eliminate.

In the work of Whittaker and Aitken, $m = 3$ was emphasized in formulating equation (1), but it was recognized that the choice of the integer m was arbitrary and that other formulations such as $m = 2$ might well be used (cf. Aitken, 1958). The parameter λ controls the magnitude of the smoothing penalty.

The form of the criterion (1) was obtained by Whittaker (1923) using Bayes rule (or inverse probability, as it was then called) using Gaussian assumptions on X^n to construct the likelihood and a conjugate Gaussian prior for S_{mn}. The solution then simply maximizes the posterior probability of a model for the data allowing for varying degrees of smoothness given λ. Whittaker's approach was therefore modern in its formulation, corresponding to Bayes and information theoretic model selection methods.

Aitken was motivated to develop and justify an algorithm that facilitated numerical calculation of the smoother. In the general case (equation (1)) for an arbitrary integer m, the operator form of the solution is, as shown in the Appendix,

$$\hat{f}_t = \frac{1}{1 + \lambda\Delta^{2m}(-L)^{-m}} X_t = \frac{L^m}{L^m + \lambda(-1)^m \Delta^{2m}} X_t, \tag{2}$$

where L is the lag operator and $\Delta = 1 - L$. Aitken gave the solution when $m = 3$, namely,

$$\hat{f}_t = \frac{1}{1 - \lambda\Delta^6 L^{-3}} X_t = \frac{L^3}{L^3 - \lambda\Delta^6} X_t, \tag{3}$$

in equation (4.2) of his (1926) paper. One of the technical contributions of this paper was to express the operator $L^3/(L^3 - \lambda\Delta^6)$ in a convergent Laurent series in powers of L and L^{-1}, so that the resulting data smoother $\hat{f}_t$ could be written as a linear combination of past and future X_t. This approach had a huge advantage over the Taylor series approximations used in earlier work since those series representations were not convergent. To simplify further and produce a numerical algorithm, Aitken solved the sixth-order polynomial equation to locate the zeros in the denominator of equation (3) and used contour integration to evaluate the coefficients in the Laurent expansion. Aitken's calculations were a major breakthrough enabling computation for much larger values of n than had previously been possible.

Nowadays, of course, computations to solve equation (1) for $\hat{f}_t$ can be done directly by matrix inversion and numerical methods. We detail the formulae and some useful new expansions in the general case here since these appear not to be in the literature. Define the following m-differencing matrix of order $(n - m) \times n$

$$D' = \begin{bmatrix} d'_m & 0 & \cdots & 0 \\ 0 & d'_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & d'_m \end{bmatrix},$$

where

$$d'_m = \left[\binom{m}{0}, (-1)\binom{m}{1}, \ldots, (-1)^{m-1}\binom{m}{m-1}, (-1)^m\binom{m}{m}\right],$$

and let $\hat{f} = (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n)'$ and $X = (X_1, X_2, \ldots X_n)'$. Then the criterion (1) has the following matrix form

$$\hat{f} = \underset{f}{\arg\min}\ \{(X - f)'(X - f) + \lambda f' D D' f\}, \tag{4}$$

whose solution is simply $\hat{f} = (I + \lambda DD')^{-1} X$. As $\lambda \to 0$, we clearly have $\hat{f} \to X$, so that the solution tracks the data exactly as is to be expected when the penalty becomes negligible.

From Lemma A in the Appendix, the solution of equation (4) may be written in alternative form as

$$\hat{f} = \left\{I - D\left\{\lambda^{-1} I + D'D\right\}^{-1} D'\right\} X,$$

and, as $\lambda \to \infty$, Lemma A gives the following asymptotic expansion in λ^{-1}

$$\hat{f} = \left\{ I - D(D'D)^{-1}D' + \lambda^{-1}D(D'D)^{-2}D' + O\left(\lambda^{-2}\right) \right\} X$$
$$= \left(I - D(D'D)^{-1}D' \right) X + O\left(\lambda^{-1}\right).$$

Then $\hat{f} \to (I - D(D'D)^{-1} D')X$, as $\lambda \to \infty$. As shown in Lemma B in the Appendix, we may write $I - D(D'D)^{-1}D' = R(R'R)^{-1}R'$, where

$$R = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^{m-1} \\ 1 & 3 & \cdots & 3^{m-1} \\ & & \ddots & \\ 1 & n & \cdots & n^{m-1} \end{bmatrix} \tag{5}$$

is an $n \times m$ orthogonal complement of the differencing matrix D. It follows that

$$\hat{f} \to R(R'R)^{-1}R'X = Ra, \quad \text{as } \lambda \to \infty. \tag{6}$$

Thus, the solution $\hat{f}$ tends asymptotically to a trend polynomial of degree $m-1$ as $\lambda \to \infty$.

When $m = 2$, the solution of equation (1) has the form

$$\hat{f}_t = \frac{1}{1 + \lambda(1-L)^4 L^{-2}} X_t = \frac{L^2}{L^2 + \lambda \Delta^4} X_t.$$

For business cycle calculations, the cyclical component is estimated as the residual $\hat{c}_t = X_t - \hat{f}_t$, leading to

$$\hat{c}_t = \frac{\lambda(1-L)^4 L^{-2}}{1 + \lambda(1-L)^4 L^{-2}} X_t = \frac{\lambda(1-L)^4}{L^2 + \lambda(1-L)^4} X_t.$$

Observe that if X_t has a stochastic trend with a unit root, then $(1-L)X_t = u_t$, for some $I(0)$ or stationary u_t. We deduce that

$$\hat{c}_t = \frac{\lambda(1-L)^3}{L^2 + \lambda(1-L)^4} u_t, \tag{7}$$

which can be expanded as a (two-sided) linear process in u_t and is stationary, confirming the trend extraction process (elimination of the unit root).

As the sample size n increases and typically as the sampling frequency increases, it is common to choose larger values of the smoothing parameter λ. For instance, the package computer program EVIEWS has automated settings of $\lambda = 100$ for annual, 1600 for quarterly, and 14, 400 for monthly data (the quarterly 1600 setting became customary following the experimentation in Hodrick & Prescott, 1997). Under such conditions, it is possible to investigate asymptotic behavior of the smoother in more detail than the result (equation (6)) given above. Note that when $(1-L)X_t = u_t$ and λ is large we have the crude approximation

$$\widehat{c}_t = \frac{(1-L)^3}{L^2/\lambda + (1-L)^4} u_t = \left\{(1-L)^{-1} + O\left(\lambda^{-1}\right)\right\} u_t \sim (1-L)^{-1} u_t,$$

so that the residuals from the smoothed series retain the unit root property. Hence, the properties of the smoother and induced 'trend' and 'cycle' are seen to be heavily influenced by the value of the smoothing parameter λ, and the smoother is not necessarily consistent for the trend component when $n, \lambda \to \infty$. Moreover, contrary to common belief based on equation (7), the business cycle estimate that is produced by HP filtering may still retain a stochastic trend asymptotically when the data follow a unit root process.

(ii) Least squares projection geometry

Aitken's (1934) article is deservedly cited for the development of generalized least squares (GLS). Little known (and hardly ever referenced) is its other major contribution: the development of the matrix algebra and projection geometry of least squares regression. Aitken initiated the matrix formulation of the regression model, the least squares criterion and its solution

$$\hat{\beta} = \arg\min_{\beta} (y - X\beta)'(y - X\beta) = (X'X)^{-1} X'y,$$

gave the fitted value $\hat{y} = Py$ and residual $\hat{u} = (I - P)y$ in terms of the projection matrix $P = X(X'X)^{-1}X'$, and discussed the properties of this projector. In doing so, Aitken provided the matrix algebra for regression that is at the heart of all modern textbook treatments and the foundation stone for the subsequent algebraic development of instrumental variable and generalized method of moment estimation.

For correlated data, where V is the variance matrix of the data, Aitken provided the corresponding GLS criterion and solution

$$\hat{\beta} = \arg\min_{\beta} (y - X\beta)' V^{-1} (y - X\beta) = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y,$$

with fitted values $\hat{y} = Py$ expressed in terms of the (non-orthogonal) projector $P = X(X'V^{-1}X)^{-1}X'V^{-1}$. Using matrix methods to minimize the trace and sums of the other principal minors of the positive definite matrix form $C'VC$, Aitken went on to prove that the GLS estimator is the best linear unbiased estimator. This elegantly written paper therefore contains the essential foodstuff of an econometrics course on linear regression. It is but seven pages long.

(iii) Minimum variance unbiased estimation

A subsequent paper by Aitken and Silverstone (1942) reports more general results on unbiased estimation with minimum sampling variance. The results were apparently obtained in Harold Silverstone's (1939) doctoral dissertation under Aitken's direction. Silverstone was the second New Zealand PhD student supervised by Aitken at Edinburgh and later wrote one of the obituaries of Aitken.

With characteristic openness, the joint paper announces that

> The starting-point, the adoption of the postulates of unbiased estimate and minimum sampling variance, and the analogies which would emerge with the theories of maximum likelihood and of linear estimation by least squares, were suggested by the senior author.

Adopting these postulates, the paper shows that

> ... simple conditions emerge under which maximum likelihood provides an estimate accurately possessing minimum variance, even though the sample is finite and the distribution of the estimating function is not normal. At the same time, the actual value of the minimum variance is obtained with special ease.

A general model is put forward in which the likelihood depends on an unknown parameter θ and the question addressed is to find conditions under which a minimum variance unbiased estimator (MVUE) exists. The method of approach is to formulate the task of optimal estimation as an extremum problem in the calculus of variations under the unbiasedness constraint. From the Euler equation for the solution of this problem, it is discovered that such an estimating function $t = t(X^n)$ exists, where X^n is the data, provided the log likelihood $\ell_n(\theta)$ satisfies an equation of the form

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = (t - \theta)/\lambda(\theta), \tag{8}$$

for the score function $\ell_n'(\theta)$, where $\lambda(\theta)$ is a Lagrange multiplier function. The authors give several examples, showing how this process works to produce an optimal estimate in some examples but in others, such as the central location of a Cauchy distribution, is intractable and cannot be resolved into the form of equation (8). In still other cases, it is shown that an optimal estimate of a certain function of θ can be obtained, in which case equation (8) is written as $\ell_n'(\theta)=(t-\tau(\theta))/\lambda(\theta)$, where $\tau(\theta)$ is the estimating function for which t is the estimator.

The authors relate their procedure to the maximum likelihood (ML) procedure, which solves the equation $\ell_n'\left(\hat{\theta}_{ml}\right) = 0$ and therefore consists essentially of equating t to $\tau(\theta)$, while ignoring $\lambda(\theta)$. They remark that

> From our standpoint, on the other hand, the existence of $\lambda(\theta)$ is fundamental; for when $\lambda(\theta)$ exists, the estimate by maximum likelihood (of $\tau(\theta)$, that is, or the trivial linear function $a\tau(\theta) + b$, but not of non-linear functions of $\tau(\theta)$) has also minimum variance, even in the case of a finite sample.

The authors go on to find (the now well-known) expression for the variance of the optimal estimate of $\tau(\theta)$, which they show to be given by

$$\left[E\left\{-\frac{\partial^2 \ell_n(\theta)}{\partial \theta^2}\right\}\right]^{-1} = i(\theta)^{-1},$$

the inverse of Fisher information. What is original to the author's treatment is that the result holds in finite samples on the condition that $\lambda(\theta)$ exists and that the score function satisfies $\ell_n'(\theta) = (t - \tau(\theta))/\lambda(\theta)$. This result was given later in Cramér's (1946) famous treatise and pre-dates the development of the so-called Cramér-Rao bound. While inequalities are not used in the Aitken–Silverstone (1942) paper, it is evident that the authors found conditions for the existence of a minimum variance unbiased estimator and found the form of its lower bound. Those conditions are shown to place the family of distributions for which a minimum variance unbiased estimator exists in the exponential family. They further show that the conditions also ensure the existence of a sufficient statistic for $\tau(\theta)$. Aitken extended this joint work to the multivariate case in a later paper (1947).

The originality of the Aitken–Silverstone paper is evident in the calculus of variations approach taken in the paper, an approach that has subsequently been used in many other contexts in statistical theory and econometrics. Its originality is also manifest in the absence of any earlier work of its type – the paper cites only two references, Fisher's (1921) paper on maximum likelihood, and Koopman's (1936) paper on exponential families. As Vere-Jones et al. (2001) put it in their biographical study of Harold Silverstone, the paper

> ... was important in its own right as a major step in the elucidation of the properties of minimum variance estimates and as a precursor to the discovery of the Cramér-Rao lower bound.

Progeny

Silverstone was one of several New Zealand students that Aitken supervised at Edinburgh. The Mathematics Genealogy Project lists his progeny as 26 doctoral students and 499 grandstudents. One of his most notable students was Henry Daniels, who originated the use of saddlepoint approximations in statistics and who was a colleague of the present author at the University of Birmingham during the 1970s. Daniels' most famous student was David Cox, one of the foremost statisticians of the second half of the 20th century. The work of both Daniels and Cox has impacted econometrics in important ways and in different fields, from structural modeling and finite sample theory to non-parametrics and from microeconometrics and time series through to financial econometrics. Through their work, Aitken's influence on econometrics has become even wider.

Rex Bergstrom

Life and career

Rex Bergstrom was born in Christchurch in 1925 and died in London in 2005. Obituaries were published in *New Zealand Economic Papers, Asymmetric Information*, and the *New Zealand Herald*. Memorial conferences were held at the University of Westminster (2005) and at the University of Essex (2006), where Bergstrom was Professor of Economics since 1970 and Emeritus Professor of Economics following his retirement in 1992. A memorial issue of *Econometric Theory* was published in 2009. The following account of Bergstrom's life and work draws from several obituaries by the present author (Phillips, 2005a,b & c) to which readers are referred for further detail.

Bergstrom's grandfather and grandmother, Sten and Sofia Bergström, had migrated to New Zealand from Sweden in 1875 and established a hotel in the town of Kumara on the West Coast of the South Island of New Zealand. Some memorabilia of the Bergstrom hotel are on exhibit at the Hokitika museum.

Bergstrom studied at Christchurch University College (1942–1947), held part-time accountancy posts and served with the RNZAF (1945–1946). His first intellectual love was mathematics and he became attracted to economic theory and the challenges of econometrics at a time when these subjects were becoming increasingly mathematical. He gained an M.Com (University of New Zealand) with first class honours in 1948 and won a Travelling Scholarship in Commerce in 1950, which he took up two years later to do his doctoral work at the University of Cambridge in 1952. His primary advisor at Cambridge was Richard Stone, whose approach to empirical research and econometric modeling was a lifelong influence.

Bergstrom had wide-ranging interests in economics, which was always apparent in conversation with him and is amply revealed in his comprehensive and lucid text (1967) on the construction of economic models. During the 1950s at the University of Auckland he taught most of the courses offered by the department of economics and, throughout his career, he regularly taught intermediate microeconomics, which was one of his favorite courses outside of econometrics. While maintaining catholic general interests in economics, Bergstrom was highly focused as a researcher. He may fairly be regarded as New Zealand's first econometrician in the modern sense of the term: someone who forges the tools that enable and empower empirical economic research. In this respect, Bergstrom was a second generation econometrician, following in the footsteps of Tinbergen, Haavelmo and Koopmans. From the remote shores of New Zealand, he joined an elite rank of researchers in the UK and the USA who were in the process of giving econometrics its own identity as a discipline by the defining nature of their work.

Bergstrom was the second New Zealander (his colleague, Bill Phillips, at the LSE in the 1960s was the first) to be elected a Fellow of the Econometric Society. Later in life, Bergstrom would joke that New Zealand had the highest number of Fellows of the Econometric Society per capita in the world. In the latter part of Bergstrom's life, the ratio of New Zealand Fellows to Members of the Econometric Society was around 50%, another remarkable statistic.

Bergstrom himself published five articles in the Econometric Society's journal *Econometrica*, the leading quantitative journal in economics. Four of these articles became landmarks that made his name in the international community and earned him professional distinction as New Zealand's leading quantitative economist.

Bergstrom's academic career began with an assistant Lectureship in economics at Massey College (1948–1949). He moved to Auckland University College in 1950 as a Junior Lecturer in a small department of four headed by Colin Simkin. Simkin had a great respect for quantitative economics and had acquired all the back issues of *Econometrica* for the departmental library as well as the University's entire collection of statistics journals. Statistics teaching at Auckland originated in the School of Commerce in 1906 with a course taught by Joseph Grossman, the founding director of the School. The ensuing tradition of statistics teaching within the economics department was substantially enhanced by Bergstrom in the 1950s and 1960s. Simkin and Bergstrom formed a powerful academic duet, making a lasting impression on the direction of economics at Auckland, becoming a strong force within the University of Auckland, and having an impact wider afield on academic economics in

Australasia. In his history of the department of economics at Auckland, Robin Court (1995), remarks of Simkin that

> He early introduced the econometric ideas of Tinbergen and Haavelmo to this part of the world, causing A W H Phillips to remark in the late 1950's that Auckland was then well ahead of most places outside the USA (and including LSE) in the range and quality of econometrics material offered.

The changes in the Auckland department were evident in its course requirements. Blyth (2004) reports that

> By 1962, the problem of mathematical background had reached the point where students were advised that before attempting stage III econometrics they should include stage I papers in pure mathematics in their degree, and before attempting Masters econometrics they should include Stage II pure mathematics.

By the end of the 1960s, Bergstrom had raised Auckland's graduate econometrics teaching to the level of the most advanced courses in leading schools in the USA and the UK, as the experiences of several of Bergstrom's students of that era who subsequently went overseas, including the present author, attest. Court's (1995) history of the Auckland department goes on to note that historically this department had

> pioneered teaching in and introduced the original courses at Auckland in a number of academic areas, including (early on) statistics, economic geography, actuarial mathematics, political economy, finance and (later) operations research ... The statistics papers in economics, first taught by Grossman from 1906, then by Neale and later by Bergstrom and Fisher, constituted the only statistics of any kind regularly taught at Auckland University College (now the University of Auckland) until 1951.

Bergstrom left Auckland on leave to do his PhD at Cambridge (1952–1954) and a decade later to take up a Readership at the London School of Economics (1962–1964) at a time when the LSE was blossoming into the leading centre of econometrics in the UK. He returned to the University of Auckland as Professor of Econometrics (1964–1971) and left New Zealand again in 1970 to become Keynes Visiting Professor at the University of Essex. He remained at Essex as Professor of Economics until his retirement in 1992, becoming the University's first Professor Emeritus of Economics. He continued academic research working from his apartment in London with undiminished vigour through to the time of his death in 2005.

Bergstrom had a profound impact on all those with whom he had contact. Small in stature and sartorial in dress in his black Saville Row suit, he brought gravitas to any academic gathering with his sharply focused intellect that uplifted the level of discussion, providing a role model for the young econometricians around him. Beyond the legacy of his writings and research, he is remembered through the Bergstrom Prize that is awarded every 2–3 years to outstanding young New Zealand econometricians, his ET Interview with the present author (Phillips, 1988), and a memorial issue of *Econometric Theory*, in which the editors described him thus:

> As a colleague he was deeply respected for upholding the highest standards of scholarship; as a teacher and supervisor he was generous, caring and inspirational; and as a person he held a quiet moral authority that guided his life and career. (Chambers, Phillips & Taylor, 2009)

Research and influence

Bergstrom's best known and most influential research contributions were to continuous time modeling, a field that he helped to establish and nurture. His other pioneering work was in exact finite sample theory, a research area that he initiated independently and at the same time as the US econometrician Robert Basmann (1961). Bergstrom also had a major influence on the development of empirical econometric research on the New Zealand economy. Each of these contributions has been reviewed in the recent obituaries of Bergstrom, so the following account simply overviews the main features of this work and explores some of its intellectual origins and its connections to ongoing research.

(i) Empirical research on the New Zealand economy

As mentioned above, when Bergstrom joined the Auckland Department of Economics in 1950, Simkin was already a strong proponent of empirical econometric research. Malcolm Fisher had completed his MA thesis, entitled 'Problems of demand measurement', in 1947 under Simkin's supervision and was acknowledged to be the department's expert in econometrics at the time that Bergstrom joined. Fisher's thesis overviewed demand analysis, the theory of identification, and regression methods, and then proceeded to conduct an empirical application to the demand for cheese in New Zealand. Blyth (2004) remarks that

> This thesis was completed inside a year, and was supervised by Simkin. It probably would have deserved a PhD at many respectable universities. It is clear that both Fisher and Simkin were well-read in the latest theoretical and statistical literature, and while there is no suggestion that the thesis made an original contribution to economic or statistical theory, in applications Fisher and his mentor were in the forefront of econometrics.

Fisher's empirical results were disappointing, reflecting the small sample of observations and the need for a complete model rather than single equation analysis. These were themes that Bergstrom was to take up in the early part of his own career, first with his doctoral research at Cambridge and later with his work on finite sample theory.

When Bergstrom arrived at Cambridge, he had already fully laid out his research topic on 'An econometric study of supply and demand for New Zealand's exports', he had collected the data needed for empirical estimation, and he had gone a long way towards formulating the econometric model that he planned to use in the empirical study. The major remaining work involved the econometric methodology (estimation procedures, the treatment of residual serial correlation, methods of prediction) and implementation, which in the early 1950s presented a major obstacle. As summarized in Chambers et al (2009), Bergstrom's PhD thesis

> ... was an extensive and original undertaking that broke new ground in empirical econometrics. The model had 27 equations and 55 parameters. It was the first large-scale macroeconometric model to be estimated by the new method of limited information maximum likelihood (LIML). Impressive in scale, design and econometric methodology, the undertaking was equally impressive in its execution, much of the calculation being done by Bergstrom's hand on an electronic desk calculator.

The thesis was supervised by J.R.N. Stone and A.D. Roy and, although Stone's own research in applied econometrics provided an important paradigm, it seems that advisor input on Bergstrom's thesis was minor. Bergstrom's study was published in *Econometrica* in 1955. It brought him international recognition, confirmed him as New Zealand's leading empirical economist and econometrician, and

> its skilful mix of economic theory, econometric methodology, and painstaking empirical implementation became the hallmark of Bergstrom scholarship. (Phillips, 2005a)

With this foundational empirical research as a beginning, it was natural to expect Bergstrom to develop into a leading applied econometrician. But Bergstrom's fascination with econometrics ran much deeper than empirical implementation. During the 1950s and 1960s he became absorbed by interests in econometric methodology and the economic theory underpinnings of empirical models. These interests led Bergstrom in two new directions.

(ii) Exact distribution theory

The first of these was an ambitious and original contribution to exact finite sample theory in econometrics published in *Econometrica* in 1962. Bergstrom's favourite book in statistics was Cramér's (1946) treatise *Mathematical methods of statistics*, which gave a particularly thorough and rigorous exposition of methods for finding the exact finite sample distribution of statistical estimators and tests. Using the methods learnt from his reading of this work, particularly the simplifying process of well-designed orthogonal transformations, Bergstrom (1962) derived the exact distributions of the maximum likelihood estimator (MLE) and the (single equation) ordinary least squares (OLS) estimator of the propensity to consume in the following simple stochastic income determination model

$$y_t = \alpha + \beta x_t + u_t \tag{9}$$

$$x_t = y_t + \gamma z_t \tag{10}$$

where the (spending propensity) parameter is β, and equation (9) has two observed endogenous variables y_t (consumption), x_t (income) and a stochastic disturbance u_t that is assumed to be *iid* $N(0, \sigma^2)$. Equation (10) is a structural (national income) identity involving an observed instrumental variable z_t that is assumed to be strictly exogenous and fixed, so that the distributional analysis is effectively conditioned on the sample $\{z_t\!: t = 1,\ldots,n\}$. The coefficient of z_t in Bergstrom (1962) was $\gamma = 1$, but it is useful to allow flexibility in γ to control the relevance of the instrument z_t in the system (Phillips, 2006). When $\gamma \to 0$, the instrument z_t becomes irrelevant to the determination of y_t and x_t, and we end up with the identity $x_t = y_t$ in place of (10). On the other hand, when $\gamma \to \infty$, the system is dominated by the signal from z_t. In view of the identity (equation (10)) and the exogeneity of z_t, the degree of endogeneity as measured by the correlation coefficient of x_t and u_t is unity, so that there is strong endogeneity in the system.

Bergstrom's expression for the exact density of the maximum likelihood estimator (in this case, indirect least squares or limited information maximum likelihood (LIML)) was easily found by transformation from the (normal) distribution of the reduced form coefficients and has the explicit form

$$\mathrm{pdf}_{\mathrm{MLE}}(b) = \frac{\lambda_n^{1/2}}{\sqrt{2\pi}} \frac{1-\beta}{\sigma} \frac{1}{(1-b)^2} \exp\left\{ -\frac{\lambda_n}{2\sigma^2} \left(\frac{b-\beta}{1-b} \right)^2 \right\}, \tag{11}$$

where $\lambda_n = \gamma^2 \sum_{t=1}^{n} z_t^2$ is a non-centrality parameter, which Bergstrom normalized to the sample size so that $\lambda_n = n$.

The exact density of the OLS estimator required substantial effort and was obtained by Bergstrom in a complicated series form but only for even values of $n \geq 4$. In recent work, the present author (Phillips, 2009) showed that this density is a specialization of a more general result (from Phillips, 1980) and has the explicit form

$$\begin{aligned} \mathrm{pdf}_{\mathrm{OLS}}(b) = & \frac{n^{\frac{n-1}{2}} \Gamma\left(\frac{n}{2}\right) e^{-\frac{n}{2\sigma^2}\left\{\frac{(b-\beta)^2}{(1-b)^2}\right\}} (1-\beta)^{n-1}}{2^{(n-1)/2} \sigma^{n-1} \sqrt{\pi} \Gamma(n-1) |1-b|^n} \\ & \times {}_1F_1\left(\frac{n}{2} - 1, n-1; -\frac{n(1-\beta)(1+\beta-2b)}{2\sigma^2 (1-b)^2} \right), \end{aligned} \tag{12}$$

involving the confluent hypergeometric function ${}_1F_1(a, b; z) = \sum_{j=0}^{\infty} \frac{(a)_j}{(b)_j j!} z^j$, where $(a)_j = \Gamma\ (a+j)/\Gamma(a)$ and $\Gamma(\cdot)$ is the gamma function. Coincidentally, this function was introduced by Whittaker, Aitken's supervisor at Edinburgh, in conjunction with several other special functions of applied mathematics, including the parabolic cylinder and Whittaker functions.

Graphs of these densities for $n = 10$ are shown in Figure 1. These were originally calculated and drawn by hand in Bergstrom (1962). The results show that the ML estimator provides an unequivocal improvement over OLS in terms of its concentration probability about the true value of β, thereby strongly vindicating the use of systems methods of estimation which take account of the structural nature of the consumption function, equation (9).

Bergstrom's paper is remarkable not only in terms of the clarity of its conclusions and the strong support that it gave to the use of simultaneous equations methodology, but also because it was Bergstrom's first piece of technical research in econometrics. Later in life, Bergstrom told me the story of how the great Norwegian econometrician Trygve Haavelmo (1989 Nobel laureate in economics) had run up to him at a European meeting of the Econometric Society in great excitement about Bergstrom's paper, telling him how he had himself worked on the same problem (the exact distribution of the OLS estimator) for ten years without success. Bergstrom told me that the derivation took him three weeks of work over a New Zealand summer.

In recent years, there has been much research on the properties of the LIML estimator in relation to least squares and instrumental variable methods. All of this

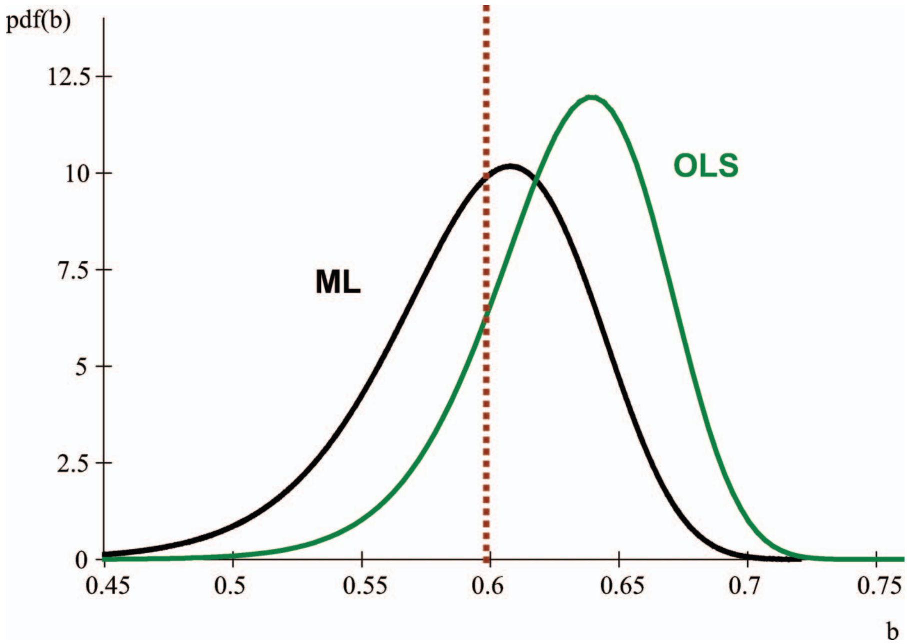


Figure 1. Densities of the maximum likelihood (ML) and least squares (OLS) estimates of the marginal propensity to consume $\beta = 0.6$ when $n = 10$.

work has reinforced Bergstrom's conclusions concerning the superiority of LIML in small samples. This conclusion remains so even though the distribution of LIML is known to have heavy tails (Phillips, 1984) and can be bimodal. The bimodality is not apparent in Figure 1 or in Bergstrom's (1962) sketch, but it becomes apparent over a wider domain, particularly when the equation is weakly identified (i.e. when γ and λ_n are small), a topic that is now of substantial interest in econometrics. Figure 2 (reproduced from Phillips, 2006) shows how the LIML distribution changes for small λ_n and how even the asymptotic theory is affected.

As is apparent from Figure 2, as λ_n decreases, the instrument z_t becomes weaker and the secondary mode in the distribution of the ML estimator becomes stronger. As explained in Phillips (2006), the income determination model (equations (9) – (10))

> is a case of strong endogeneity, where there is a structural behavioral equation and an identity. The identity is another structural relation and its role is important in the distribution theory because it provides a magnet for an alternative centering, pulling consistent estimators like IV and LIML away from the relevant parameter in the behavioral relation and thereby naturally inducing a bimodality. In fact, it is the identity that is the source of the bimodality.

The coefficient in the structural identity provides a point of compression in the density that gives rise to the bimodality. Weak instrumentation is an energy source that feeds the secondary mode in the distribution on the right side of unity. The process works like a compressed balloon, which upon inflation expands on either side of the point of compression. As $\lambda_n \to 0$, the limit distribution ends up being symmetric about unity (the coefficient in the identity) with equal modes on either side.

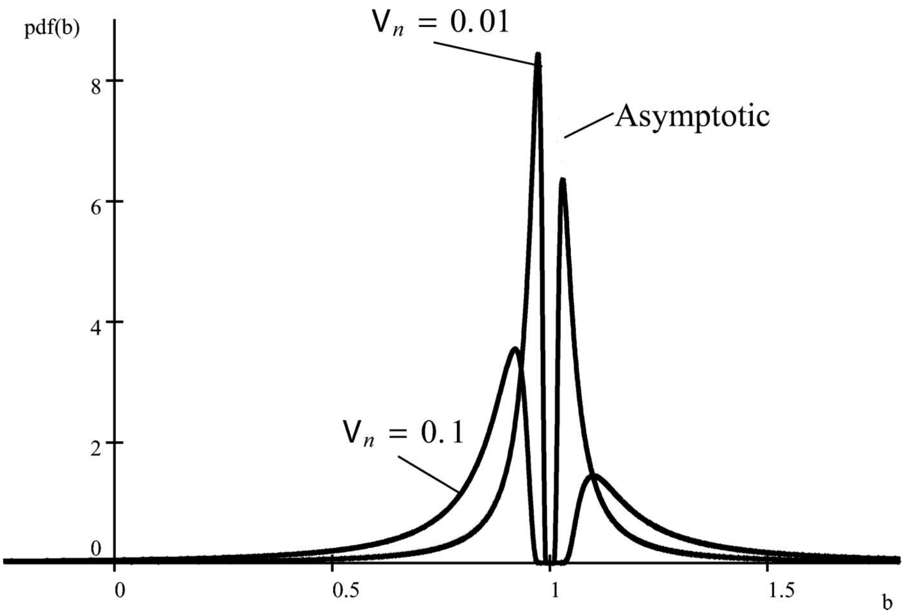


Figure 2. Densities of the ML estimator of $\beta = 0.6$ for $\lambda_n = 0.1$, 0.01, and the (very weak) instrument limit distribution of ML (circled).

(iii) Continuous time econometric models

Bergstrom's focus in his pursuit of continuous time econometric modeling was macroeconomic activity. His interest was stimulated by the early research of fellow New Zealander Bill Phillips, whose work he greatly admired. During the 1950s, Phillips had worked at the LSE on continuous time formulations of trade cycle models and policy adjustment mechanisms. In the early 1960s, this work was extended to accommodate cyclical growth. Phillips (1959) had also attempted to develop a methodology for estimating such continuous systems with discrete data in a paper published in *Biometrika*. Earlier arguments for the use of continuous time dependencies in econometric modeling had been proposed by Koopmans (1950). By the mid 1960s, Bergstrom had become convinced that such formulations were the most effective way of introducing economic theory restrictions into econometric specifications and to allow for the reality that aggregate economic activity depends on the continuous passage of time. In effect, the economy does not cease to exist between discrete measurements.

To Bergstrom, these arguments were compelling and he fell under the grip of an intellectual fascination that would overtake the rest of his life. By the early 1970s, no conversation with Bergstrom was complete without some mention of continuous time modeling and the latest results and problems in the field. As the years went by the topics of conversation evolved with the development of the subject: the findings from the empirical models, the problems presented by various data types, the formulation of higher order models, algorithms for constructing the likelihood, and the implications of stochastic trends. The field presented a vast arena for new research. As each problem was solved, new problems arose in their place. The program of research took on the aspect of an enterprise resembling that of the Cowles Commission researches in the 1940s, but instead of a vast team of researchers, the endeavour was led by a single man armed only with his own vision and successive generations of students who worked under him, aided by the occasional assist from an interested academic interloper. By the year 2000, a huge amount of theoretical and algorithmic work had been accomplished, several generations of empirical models of the UK had been constructed, estimated and tested, and continuous time models had been estimated and were in use by central banks for several OECD countries.

Bergstrom's work in this field has been reviewed in detail in the obituaries cited earlier, in two of his own overviews (Bergstrom, 1988, 1996) and, most recently, by Bergstrom's former student and co-author Nowman (2009). At the time of Bergstrom's death, he had just completed a monograph written jointly with Nowman, *A continuous time econometric model of the United Kingdom with stochastic trends*, which developed, estimated and implemented the latest version of his continuous time econometric model of the UK economy. In a Foreword to this volume (Bergstrom & Nowman, 2006), the present author wrote that the work

> carries the indelible signature of Bergstrom's superb scholarship. The theoretical model is developed with great attention to underlying economic ideas, the econometric methodology is systematically built on the extraction of an exact discrete model and an algorithm that constructs the Gaussian likelihood, the empirical implementation is painstakingly conducted, the size of the system and its complex transcendental matrix nonlinearities push to the limits of present computational capacity, and the empirics involve specification testing and prediction evaluation against a highly competitive VAR system with exogenous inputs.

As Chambers, Phillips and Taylor (2009) put it in the memorial volume to Bergstrom:

> Perhaps uniquely in our profession Rex dedicated the last 40 years of his working life to pursuing a single research goal – the methodological development of econometric models in continuous time and the empirical application of these models to macroeconomic activity. This long term project evolved into the centrepiece of his life as a researcher.

No one can doubt the enormous personal commitment that Bergstrom made to his central research goal of developing continuous time econometrics. But the impact of his work in macroeconomics has been less than he hoped. In macroeconomic theory, his approach was perceived as being outside the mainstream concerns of dynamic macro, where small-scale intertemporal optimization models have held centre stage for three decades. In the world of macroeconometric modeling, agnostic time series and multivariate analysis methods have retained their popularity in empirical applications. No doubt, convenience is a major factor. The Bergstrom approach requires mastery of much theory and econometric methodology on the part of an empirical research team, implementation can present formidable numerical challenges, and simple changes in specification can involve major systemwide implications. To Bergstrom and those trained in the approach, these challenges constitute the essence of good empirical modeling.

In financial econometrics, the methods of continuous time econometrics have been truly vindicated. The last decade has seen ultra high frequency datasets become standard with electronic monitoring of financial transactions. Continuous time stochastic process methods are the central developmental tool in financial theory.

Parametric econometric modeling in continuous time using maximum likelihood or approximate maximum likelihood methods is a gold standard in empirical research. And non-parametric approaches to volatility measurement using empirical (discrete) versions of continuous quadratic variation processes has opened up many new research avenues and affected industry practice. All of these developments were welcomed by Bergstrom and fit comfortably within his mantra of continuous time modeling and empirical practice.

Progeny

The Mathematics Genealogy Project lists Bergstrom's progeny as 171 descendants, which underestimates the actual total because most of his PhD students and their students are not presently listed in the mathematics genealogy record. At the University of Auckland, Bergstrom supervised the department's first two methodological masters theses (MacCormick, 1969; Phillips, 1970) and the department's first PhD in economics (Hall, 1971). At the University of Essex, Bergstrom supervised nine PhD students over the period 1971–1993, among whom were his co-authors Marcus Chambers and Ben Nowman. A complete list of the students whose theses were supervised by Bergstrom is given in Phillips (2005a). Chambers, Nowman and several of their students have actively pursued Bergstrom's research on continuous time modeling. Some have continued his work on exact distribution theory. Many have become macroeconomists and empirical researchers. Others have moved into the world of business and the finance industry.

Bergstrom had a profound impact on his students. His scholarship, commitment to research, and personal integrity were an attractive force and enduring example to all around him. His lectures opened up new worlds of technical understanding and insight to successive generations of economics students at Auckland and Essex. They were

> models of clarity. He came to lectures extremely well-prepared, carrying a folder of notes that he usually left unopened on the front desk, presumably for reference should he need it. He then proceeded to develop all the material on the blackboard with great precision and economy of presentation without consulting his notes and usually with little class interaction, students simply watching in wonderment at what transpired at the board. (Phillips, 2005a).

Conclusion

Alec Aitken began his research before the official birth of econometrics, which can conveniently be dated to the foundation of the Econometric Society in 1930 and the establishment of the journal *Econometrica*, which published its first issue in 1933. By the time he retired in 1965, a second generation of econometrician was in full flight. Within that generation, Rex Bergstrom was a significant figure.

Prior to the formation of the Econometric Society, earlier generations of economists had made considerable contributions to statistical method, Francis Ysidro Edgeworth and Irving Fisher being two prominent examples. Aitken's work was distinguished because it helped to build the mechanics of modern econometrics with smoothing algorithms, the matrix mechanics of linear regression, and the statistical properties of estimation. Bergstrom took up the mantle of building the methodology of econometrics in the generation that followed. He promoted an agenda of research that was also notable for its mathematical rigour, but emphasized its connection with economic theory, and its practical relevance to empirical econometric modeling. Together, the work of Aitken and Bergstrom laid some early foundations for which subsequent generations of econometrician are indebted and from which a substantial edifice of theory and applications continues to grow.

New Zealand economists have established a notable tradition in econometrics. Bergstrom is justly known and honoured as a significant pioneer in establishing that tradition. Aitken, on the other hand, is virtually unknown in the world of economics and little known among econometricians. While seldom cited, his contributions to linear regression, minimum variance unbiased estimation, and data smoothing algorithms are now part of the fabric of modern econometric practice. These contributions earn Aitken a significant place in the history of econometrics and make him one of New Zealand's pioneers in this field.

Acknowledgements

This paper was presented in July 2009 at the New Zealand Association of Economists Conference, Wellington. Thanks are due to Olav Bjerkholt and Deborah Blood for helpful comments on the original version. This research was supported by the NSF under Grant No. SES 06–47086.

Notes

1. The matrix treatment of correlation and scatter analysis was originally promoted by Frisch (1929).
2. The Guiness book of records and internet list some astonishing recently recorded feats of memory and mental arithmetic of the same genre as those demonstrated by Aitken. Daniel Tammet (an autistic savant) is said to have recited π to 22,514 digits in 5 hours 9 minutes; Shakuntala Devi, known as the human computer in India, multiplied two 13 random digit numbers (picked by the computer science department at Imperial College) in 28 seconds in 1980 (a feat mentioned in the *Guiness book of records*, 1995, p. 26). One of the famous Chinese Kaohsiung siblings, Wang Chia-lu, multiplied two 13 digit numbers in 26.51 seconds in 2000. Alexis Lemaire (who, at the time of writing, is a PhD student in artificial intelligence) found the 13th root of a random 200 digit number in 70 seconds at the London Science Museum.
3. This paper is usually cited as Whittaker (1923), but in the journal records on Cambridge Online Vol. 41 is listed as appearing in February 1922. At the head of the article itself, it is recorded that it was 'received in amended form' in August 1923.

References

Abadir, K., & Magnus, J.R. (2005). *Matrix algebra*. *Econometric exercises* (Vol. 1). Cambridge: Cambridge University Press.

Aitken, A.C. (1925). *The smoothing of data*. DSc dissertation, University of Edinburgh.

Aitken, A.C. (1925). On Bernoulli's numerical solution of algebraic equations. *Proceedings of the Royal Society of Edinburgh*, *46*, 289–305.

Aitken, A.C. (1926). On the theory of graduation. *Proceedings of the Royal Society of Edinburgh*, *47*, 36–45.

Aitken, A.C. (1931). On interpolation by iteration of proportional parts without the use of differences. *Proceedings of Edinburgh Mathematical Society*, *3*, 56–76.

Aitken, A.C. (1934). On least squares and linear combinations of observations. *Proceedings of the Royal Society of Edinburgh*, *55*, 42–48.

Aitken, A.C. (1937). Trial and error and approximation in arithmetic. *Mathematical Gazette*, *21*, 117–122.

Aitken, A.C. (1939a). *Determinants and matrices*. London: Oliver and Boyd. (Many later editions.).

Aitken, A.C. (1939b). *Statistical mathematics*. London: Oliver and Boyd. (Many later editions.).
Aitken, A.C. (1947). On the estimation of many statistical parameters. *Proceedings of the Royal Society of Edinburgh Section A*, *55*, 369–377.
Aitken, A.C. (1958). The contributions of E. T. Whittaker to Algebra and numerical analysis. *Proceedings of the Edinburgh Mathematical Association*, *11*, 31–38.
Aitken, A.C. (1954). The art of mental calculation: with demonstrations. *Transactions of the Royal Society of Engineers*, London, *44*, 295–309.
Aitken, A.C. (1962). *The case against decimilization*. London: Oliver and Boyd.
Aitken, A.C. (1963). *From Gallipoli to the Somme: Recollections of a New Zealand infantry man*. London: Oxford University Press.
Aitken, A.C., & Silverstone, H. (1942). On the estimation of statistical parameters. *Proceedings of the Royal Society of Edinburgh*, *A61*, 186–194.
Aitken, A.C., & Turnbull, H.W. (1932). *An introduction to the theory of canonical matrices*. Edinburgh: Blackie (with later editions).
Basmann, R.L. (1961). Note on the exact finite sample frequency functions of generalized classical linear estimators in two leading overidentified cases. *Journal of the American Statistical Association*, *56*, 619–636.
Bergstrom, A.R. (1955). An econometric study of supply and demand for New Zealand's exports. *Econometrica*, *23*, 258–276.
Bergstrom, A.R. (1962). The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. *Econometrica*, *30*, 480–490.
Bergstrom, A.R. (1967). *The construction and use of economic models*. London: English Universities Press.
Bergstrom, A.R. (1988). The history of continuous-time econometric models. *Econometric Theory*, *4*, 365–383.
Bergstrom, A.R. (1996). Survey of continuous time econometrics. In W.A. Barnett, G. Gandolfo, & C. Hillinger (Eds.), *Dynamic disequilibrium modelling* (pp. 3–25). Cambridge: Cambridge University Press.
Bergstrom, A.R., & Nowman, B.N. (2006). *A continuous time econometric model of the United Kingdom with stochastic trends*. Cambridge: Cambridge University Press.
Blyth, C.A. (2004). A history of the Economics Department of the University of Auckland. Working Paper, University of Auckland.
Chambers, M., Phillips, P.C.B., & Taylor, A.M.R. (2009). Econometric theory memorial to Albert Rex Bergstrom. *Econometric Theory*, *25*(4), 891–900.
Court, R.H. (1995). Economics at Auckland – an outline history. Working Paper, University of Auckland.
Cramér, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.
David, H.A. (1995). Occurrence of common terms in mathematical statistics. *American Statistician*, *49*, 121–133.
Fisher, R.A. (1921). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of Royal Society of London,* Series A, *222*, 309–368.
Frisch, R. (1929). Correlation and scatter in statistical variables. *Nordixk Statistisk Tidskrift*, *8*, 36–102.
Gradshteyn, I.S., & Ryzhik, I.M. (2000). *Tables of integrals, series and products*. New York: Academic Press.
Hall, V.B. (1971). *A model of New Zealand's post-war inflation*. Thesis (PhD–Economics), University of Auckland.
Henderson, R. (1924). A new method of graduation. *Transactions of the Actuarial Society of America*, *25*, 29–40.
Hodrick, R.J., & Prescott, E.C. (1997). Postwar business cycles: an empirical investigation. *Journal of Money Credit and Banking*, *29*, 1–16.
Kidson, H.P. (1968). Alexander Craig Aitken, 1895–1967: notes toward a biography. *Comment: A New Zealand Quarterly Review*, April, 12–19.
Koopman, B.O. (1936). On distributions admitting a sufficient statistic. *Transactions of the American Mathematical Society*, *39*, 399–409.
Koopmans, T.C. (1937). *Linear regression analysis of economic time series*. Haarlem: De Erven F. Bohn.

Koopmans, T. (1950). *Statistical inference in dynamic economic models*, Cowles Commission Monograph No. 10. New York: Wiley.
Leser, C.E.V. (1961). A simple method of trend construction. *Journal of the Royal Statistical Society, Series B*, *23*, 91–107.
MacCormick, A. (1969). *Regulation of economic systems by linear least-squares methods*. Thesis (M.Com), University of Auckland.
McCloskey, D. (2000). *Economical writing*. Prospect Heights, IL: Waveland Press Inc.
Nowman, K.M. (2009). Rex Bergstrom's contributions to continuous time macroeconometric modeling. *Econometric Theory*, *25*, 301–327.
Phillips, A.W.H. (1959). The estimation of parameters in systems of stochastic differential equations. *Biometrika*, *46*, 67–76.
Phillips, P.C.B. (1970). *The structural estimation of stochastic differential equation systems*. Thesis (MA), University of Auckland.
Phillips, P.C.B. (1980). The exact finite sample density of instrumental variable estimators in an equation with $n + 1$ endogenous variables. *Econometrica*, *48*(4), 861–878.
Phillips, P.C.B. (1984). The exact distribution of LIML: I. *International Economic Review*, *25*, 249–261.
Phillips, P.C.B. (1988). The ET interview: A.R. Bergstrom. *Econometric Theory*, *4*, 301–327.
Phillips, P.C.B. (2005a). Albert Rex Bergstrom 1925–2005. *New Zealand Economic Papers*, *39*, 129–152.
Phillips, P.C.B. (2005b). Albert Rex Bergstrom: pioneer of economic modeling. *New Zealand Herald*, May 21.
Phillips, P.C.B. (2005c). Albert Rex Bergstrom: pioneer of continuous time economic models. *Asymmetric Information*, July, 3–4.
Phillips, P.C.B. (2006). A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory*, *22*(5), 947–960.
Phillips, P.C.B. (2009). Exact distribution theory in structural estimation with an identity. *Econometric Theory*, 958–984.
Royal Society Library and Archive Catalogue Citation. *Alexander Craig Aitken*, Ref# EC/1936/01.
Silverstone, H. (1939). *On the theory of estimation of statistical parameters*. Ph.D dissertation, University of Edinburgh.
Silverstone, H. (1968). Obituary: Alexander Craig Aitken 1895–1967. *Journal of the Royal Statistical Society, Series A*, *131*, 259–261.
Tee, G.J. (1981). Two New Zealand mathematicians. In J.M. Crossley (Ed.), *First Australian conference on the history of mathematics*, 180–199. Monash University.
Vere-Jones, D., MacKinnon, M.J., & Silverstone, B. (2001). Harold Silverstone: a perspective. *Australian and New Zealand Journal of Statistics*, *43*, 393–398.
Whittaker, E.T., & Watson, G.N. (1902). *A course of modern analysis*. Cambridge: Cambridge University Press.
Whittaker, E.T. (1923). On a new method of graduation. *Proceedings of the Edinburgh Mathematical Association*, *78*, 81–89.
Whittaker, E.T., & Robinson, G. (1924). *Calculus of observations. An introduction to numerical mathematics*. London: Blackie.
Whittaker, J.M., & Bartlett, M.S. (1968). Alexander Craig Aitken: 1895–1967. *Biographical memoirs of the Royal Society*, *14*, 1–14.

Appendix

Proof of equation (2)

In the general case

$$\hat{f}_t = \arg\min_{ft} \left\{ \sum_{s=1}^{n} (X_s - f_s)^2 + \lambda \sum_{s=1}^{n} \left[\sum_{j=0}^{m} \binom{m}{j} (-L)^j f_s \right]^2 \right\},$$

for which the first order conditions are:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial f_t}\sum_{s=1}^{n}(X_s - f_s)^2 + \lambda\frac{\partial}{\partial f_t}\left\{\sum_{s=t}^{t+m}(\Delta^m f_s)^2\right\} \\
&= -2(X_t - f_t) + 2\lambda\left\{\sum_{j=0}^{m}\binom{m}{j}(-1)^j\left(\Delta^m f_{t+j}\right)\right\} \\
&= -2(X_t - f_t) + 2\lambda\left\{\sum_{j=0}^{m}\binom{m}{j}\left(-L^{-1}\right)^j(\Delta^m f_t)\right\}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
X_t &= \left\{1 + \lambda\sum_{j=0}^{m}\binom{m}{j}(-1)^j L^{-j}\Delta^m\right\} f_t = \left\{1 + \lambda\left[-L^{-1} + 1\right]^m \Delta^m\right\} f_t \\
&= \left\{1 + \lambda[1 - L]^m\left(-L^{-1}\right)^m \Delta^m\right\} f_t = \left\{1 + \lambda(1 - L)^m(-L)^{-m}\Delta^m\right\} f_t.
\end{aligned}
$$

Thus

$$
\hat{f}_t = \frac{1}{1 + \lambda\Delta^{2m}(-L)^{-m}} X_t = \frac{L^m}{L^m + \lambda(-1)^m\Delta^{2m}} X_t.
$$

Lemma A. *Writing* $\hat{f} = (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_n)'$ and $X = (X_1, X_2, \ldots X_n)'$, *the solution to (1) and (4) is*

$$
\hat{f} = (I + \lambda DD')^{-1} X \tag{13}
$$

$$
= \left\{I - D\left\{\lambda^{-1} I + D'D\right\}^{-1} D'\right\} X \tag{14}
$$

$$
= \left\{I - D(D'D)^{-1}D' + \lambda^{-1}D(D'D)^{-1}D' + O\left(\lambda^{-2}\right)\right\} X, \tag{15}
$$

so that $\hat{f}_t = X_t + O(\lambda)$ *as* $\lambda \to 0$ *and* $\hat{f} = (I - D(D'D)^{-1}D'\}X + O\,(\lambda^{-1})$ *as* $\lambda \to \infty$.

Proof First-order conditions for equation (4) yield equation (13) directly. To show (equations 14) and (15) we use the inversion formula (e.g. Abadir & Magnus, 2007, 107)

$$
\left(A - BE^{-1}C\right)^{-1} = A^{-1} + A^{-1}B\left\{E - CA^{-1}B\right\}^{-1} CA^{-1},
$$

and setting $A = 1$, E=-$\lambda^{-1}I$, $B = D$, $C = D'$, we have

$$
\begin{aligned}
(I + \lambda DD')^{-1} &= I + D\left\{(-\lambda I)^{-1} - D'D\right\}^{-1} D' \\
&= I - D\left\{\lambda^{-1} I + D'D\right\}^{-1} D' \\
&= I - D(D'D)^{-1/2}\left\{I + \lambda^{-1}(D'D)^{-1}\right\}^{-1}(D'D)^{-1/2}D' \\
&= I - D(D'D)^{-1/2}\left\{I - \lambda^{-1}(D'D)^{-1} + O\left(\lambda^{-2}\right)\right\}(D'D)^{-1/2}D' \\
&= I - D(D'D)^{-1}D' + \lambda^{-1}D(D'D)^{-2}D' + O\left(\lambda^{-2}\right),
\end{aligned}
$$

giving the stated results.

Lemma B

$$I - D(D'D)^{-1}D' = R(R'R)^{-1}R' \tag{16}$$

where R is given by equation (5).

Proof Let $C = [D,R]$ and note that $D'R = 0$ because the $(t,\ p+1)$'th element of $D'R$ for $1 \leq t \leq n-m$ and $0 \leq p \leq m-1$ is

$$\begin{aligned}\sum_{j=0}^{m}\binom{m}{j}(-1)^j(t+j)^p &= \sum_{j=0}^{m}\binom{m}{j}(-1)^j\sum_{k=0}^{p}\binom{p}{k}t^{p-k}j^k \\ &= \sum_{k=0}^{p}\binom{p}{k}t^{p-k}\sum_{j=0}^{m}\binom{m}{j}(-1)^j j^k \\ &= 0\end{aligned}$$

since $\sum_{j=0}^{m}\binom{m}{j}(-1)^j j^k = 0$ for $0 \leq k \leq p < m$, as can be shown by recursion (see also Gradshteyn & Ryzhik, 2000, 0.154.3). Next observe that $H = C(C'C)^{-1/2} = [D(D'D)^{-1/2}, R(R'R)^{-1/2}]$ is orthogonal, and the stated result (equation (16)) follows.