UNIVERSITY OF OXFORD

Discussion Papers in Economic and Social History

Number 41, July 2001

A PLEA FOR ERRORS

CHARLES H. FEINSTEIN

and

MARK THOMAS

UNIVERSITY OF OXFORD

Discussion Papers in Economic and Social History

1 Hans-Joachim Voth and Tim Leunig, *Did Smallpox Reduce Height? Stature and the Standard of Living in London, 1770–1873* (Nov. 1995)

2 Liam Brunt, *Turning Water into Wine — New Methods of Calculating Farm Output and New Insights into Rising Crop Yields during the Agricultural Revolution* (Dec. 1995)

3 Avner Offer, *Between the Gift and the Market: the Economy of Regard* (Jan. 1996)

4 Philip Grover, *The Stroudwater Canal Company and its Role in the Mechanisation of the Gloucestershire Woollen Industry, 1779–1840* (March 1996)

5 Paul A. David, *Real Income and Economic Welfare Growth in the Early Republic or, Another Try at Getting the American Story Straight* (March 1996)

6 Hans-Joachim Voth, *How Long was the Working Day in London in the 1750s? Evidence from the Courtroom* (April 1996)

7 James Foreman-Peck, *'Technological Lock-in' and the Power Source for the Motor Car* (May 1996)

8 Hans-Joachim Voth, *Labour Supply Decisions and Consumer Durables During the Industrial Revolution* (June 1996)

9 Charles Feinstein, *Conjectures and Contrivances: Economic Growth and the Standard of Living in Britain During the Industrial Revolution* (July 1996)

10 Wayne Graham, *The Randlord's Bubble: South African Gold Mines and Stock Market Manipulation* (August 1996)

11 Avner Offer, *The American Automobile Frenzy of the 1950s* (Dec. 1996)

12 David M. Engstrom, *The Economic Determinants of Ethnic Segregation in Post-War Britain* (Jan. 1997)

13 Norbert Paddags, *The German Railways — The Economic and Political Feasibility of Fiscal Reforms During the Inflation of the Early 1920s* (Feb. 1997)

14 Cristiano A. Ristuccia, *1935 Sanctions against Italy: Would Coal and Crude Oil have made a Difference*? (March 1997)

15 Tom Nicholas, *Businessmen and Land Purchase in Late Nineteenth Century England* (April 1997)

16 Ed Butchart, *Unemployment and Non-Employment in Interwar Britain* (May 1997)

17 Ilana Krausman Ben-Amos, *Human Bonding: Parents and their Offspring in Early Modern England* (June 1997)

18 Dan H. Andersen and Hans-Joachim Voth, *The Grapes of War: Neutrality and Mediterranean Shipping under the Danish Flag, 1750–1802* (Sept. 1997)

19 Liam Brunt, *Nature or Nurture? Explaining English Wheat Yields in the Agricultural Revolution* (Oct. 1997)

20 Paul A. David, *Path Dependence and the Quest for Historical Economics: One More Chorus of the Ballad of QWERTY* (Nov. 1997)

21 Hans-Joachim Voth, *Time and Work in Eighteenth-Century London* (Dec. 1997)

22 Tim Leunig, *New Answers to Old Questions: Transport Costs and The Slow Adoption of Ring Spinning in Lancashire* (Feb. 1998)

23 Paul A. David, *From Keeping 'Nature's Secrets' to the Institutionalization of 'Open Science'* (July 2001)

24 Federico Varese and Meir Yaish, *Altruism: The Importance of Being Asked. The Rescue of Jews in Nazi Europe* (May 1998)

25 Avner Offer, *Epidemics of Abundance: Overeating and Slimming in the USA and Britain since the 1950s* (Nov. 1998)

[Continued inside the back cover]

A Plea for Errors

CHARLES H. FEINSTEIN

All Souls College, Oxford

MARK THOMAS

Nuffield College, Oxford

We wish to thank Hans-Joachim Voth
and Martin Weale for comments on earlier drafts

ABSTRACT

This paper argues that all historical data series should be accompanied by formal estimates of their margins of error. We discuss the nature of errors in data series and review earlier attempts to assess their reliability. We show how overall margins of error may be calculated for historical series from judgments on the reliability of their components, and how these allow readers both to appraise the estimate and to test the implications of applying different standards. An illustration is provided for Hoffmann's index of British industrial output, 1770–1831. The calculations emphasize the value of this approach to the recent debate on growth rates during the industrial revolution and suggest its merits more generally.

A Plea for Errors

The twin pillars of the modern approach to economic history are quantification and the explicit use of economic theory. Neither is new, but both have received much greater emphasis since the emergence of the 'new' economic history (or cliometrics). Measurement and theory have in turn been severely criticised by exponents of more traditional approaches, and charged with a variety of serious offences, including the introduction of irrelevant theory and the use of inaccurate data. Our concern in the present paper is with one specific aspect of the contribution of measurement to economic history: the problems of reliability that are inevitable when working with historical statistics.

These problems of measurement are sharply illustrated by the current controversy about the pace of growth during the industrial revolution. Revised estimates of the rate of growth of the British economy in the late eighteenth and early nineteenth century are at the root of the issues on which the opposing forces have taken up positions. On one side, the authors of the series present their estimates as the basis for a radically different interpretation of the nature and timing of the industrial revolution. On the other, their critics stress the unreliability of the data underlying these estimates, and assert the inability of such quantitative series to capture the full impact of the massive social and economic transformation in this period.[1]

While such disagreements can never be entirely eliminated, their scope could be greatly minimised, and the real points at issue sensibly clarified, if quantitative estimates were invariably accompanied by explicit quantitative indications of their reliability. Our plea is thus that any new statistical series should always be presented together with a careful guide to the associated margins of error. If this became standard practice, it would ensure that both authors and other users were properly aware of the extent to which they could have confidence in the estimates, and how strongly they could press any conclusions derived from these series. Similarly, it would not be good enough for critics of the analysis merely to demonstrate that the numbers were not wholly reliable. It would be incumbent on them to argue that the probable errors were greater than the compilers had claimed, and that this additional degree of error was sufficient to invalidate the conclusions. The procedure has two further advantages: it would

[1] Hoppit, "Counting;" Crafts and Harley, "Output Growth;" Berg and Hudson, "Rehabilitating the Industrial Revolution;" Jackson, "Rates of Industrial Growth;" Cuenca Esteban, "British Textile Prices" and "Rising Share of British Industrial Exports;" Harley and Crafts, "Cotton Textiles;" and Temin, "Two Views."

tend to reduce continuous iterations towards greater refinement in the numbers, if it could be shown that further revisions fall well within the margins of error of the original data; finally, in cases where the margin of error remains uncomfortably large, the procedure can direct researchers to those components in a series where more information will have the greatest effect in reducing the margin around the final estimate.

I
Historical applications of this approach

We are not, of course, the first to recommend that quantitative estimates should always be given together with an indication of their probable errors. The great pioneer in this field was Arthur Bowley. As befits a Cambridge mathematical Wrangler, he was aware of the statistical problem of errors and applied ‘the theory of error’ from the very start of his research into the changes in wages.[2] In a paper of 1911 he set out a full treatment based explicitly on a recent proof of the law of errors by Francis Edgeworth. He covered all the critical issues, including the effects of correlation within and between the constituent estimates, and extended the procedure to cover not only single values, but also weighted averages and ratios of weighted averages. At about the same time he demonstrated the application of his method in a British Association report on the ‘small incomes’ received by self-employed and salaried workers.[3]

A simpler version of this approach was adopted immediately after the war, in a study of the national income of the United States. For each item, each of the authors made a subjective estimate of ‘a range within which he thought the truth was equally likely to lie or not to lie.’ The probable errors in the aggregate were then computed ‘in the usual manner’ by squaring these estimated errors, adding the squares and extracting the square root of the sum.[4]

Unfortunately, after this very promising beginning, evaluation of errors in economic statistics faded. Bowley reproduced his main techniques in successive editions of his statistics textbook, and it was used in very summary form in his

[2] Bowley, “Changes in Average Wages,” pp. 236–7; idem, “Agricultural Wages,” p. 559.

[3] Idem, “Measurement;” British Association, “Report,” pp. 46–7 and 62–4.

[4] Mitchell, et al, *Income in the United States*, pp. 60–62.

work with Lord Stamp on national income.[5] However, it was not pursued in the subsequent studies by Colin Clark; and we have not found any serious discussion in any of the other major contributions to economic statistics in the inter-war period in either Europe or the United States. For example, no attempt was made to attach margins of error to various indices of industrial production.[6]

Interest in the subject then revived on both sides of the Atlantic, and subjective assessment of margins of error flourished for a few years during the war and early post-war period. In Britain, the leading exponent of this procedure was Richard Stone. He raised the issue in the context of his work with James Meade and David Champernowne on the wartime national accounts, and it was taken forward by a number of those who worked with him.[7] Continuity with previous practice was made explicit by R. C. Desai, a doctoral student working with Stone at Cambridge. He opened his discussion of the reliability of his estimates of consumers' expenditure in India in the 1930s by stating: 'Right from the inception of the work I have persevered with the idea of following Bowley's precept of keeping a running account of error.' Desai then noted that this required first ascertaining the error in each element in the total, and then evaluating the incidence of these errors on the total. For the latter purpose he used a more elegant formulation of the error function involving the use of calculus, a technique which he credited to the American statistician and management expert, W. Edwards Deming.[8]

Stone was also associated with the study of national incomes in three British colonies published by Phyllis Deane in 1948.[9] She too estimated the subjective margins of error of each component of her national totals, and arrived at the margins of error of the totals by what she called the 'usual procedures', indicating that the method was by then thought to be well established in the United Kingdom. As with Desai, where the errors in the components were believed to be independent, the margins of error were combined by taking the square root of the sums of the variances; where it was believed that the errors were perfectly

---

[5] Bowley, *Elements of Statistics*, pp. 312–42 and 446–9; Bowley and Stamp, *National Income*, p. 47.

[6] Clark, *National Income*, *National Income and Outlay*; Rowe, "Physical Volume of Production;" Hoffmann, *British Industry*; Fabricant, *Output*.

[7] Stone, et al "Precision;" Stone, National Income, p. 13.

[8] Desai, "Consumer Expenditure," pp. 271–3; Deming, *Statistical Adjustment of Data*.

[9] Deane, *Colonial National Income*. Miss Deane confirmed Stone's role in this application of the procedure during a conversation with Feinstein in October 1997.

correlated the margins of error were simply added together. In other cases an intermediate figure was taken. A very similar procedure was then followed by Agatha Chapman for her estimates of interwar wages and salaries in the United Kingdom and in subsequent volumes of the *Studies in the National Income and Expenditure of the United Kingdom*, for which Stone was the general editor.[10]

Further consideration was given to the theoretical aspects of the calculation of errors by Andrew Roy, an economic statistician also working in Cambridge. He commended the use of the Stone-Desai-Chapman procedure, and explored both theoretically and empirically the crucial issue of whether or not the probable errors around any given estimate can be assumed to be symmetrical. He showed that under certain assumptions the errors could be so large that the labour of making the estimates is 'found to have borne no fruit at all', and appealed for more work to be done on the possible interpretation of the errors that occur in economic statistics.[11]

In the United States, Simon Kuznets also pursued the problems of reliability in his monumental study of interwar national income, but in a different and less satisfactory manner. He and his collaborators assigned subjective margins of error to each of the components of the national income. However, they then calculated the overall error in the total by simply taking a weighted average of the component errors, so obtaining an overall margin of error of about 20 per cent.[12] This approach drew the fire of reviewers in the United Kingdom. Bowley commented caustically: 'if this statement is taken literally, it suggests that the enormous labour spent in producing this book was largely wasted.'[13] Stone was critical of several aspects of the procedure; in particular, he argued that Kuznets's failure to assign offsetting random errors to the components of national income was 'likely to result in a very considerable over-estimate of the margin of error of the total.'[14]

Indeed, Kuznets himself realised that his approach exaggerated the margins of error, but seemed unaware of the more appropriate procedures which had been developed in the United Kingdom. To illustrate the difference it would have made to the presentation and appraisal of his final estimates, we can apply

---

[10] Chapman, *Wages and Salaries*, pp. 230–6; Prest and Adams, *Consumers' Expenditure*, pp. 179–82; Stone and Rowe, *Consumers' Expenditure*, II, pp. 115–9.

[11] Roy, "Exercise in Errors," p. 514.

[12] Kuznets, *National Income*, II, pp. 501–37.

[13] Bowley, "National Income in America," p. 232.

[14] Stone, "Two Studies," p. 69.

the main elements of the Stone-Desai-Chapman procedure to the Kuznets series for total net income originating. The margin of error calculated for this item by Kuznets for 1929 to 1935 was 18.6 per cent. His method assumes that the errors in each estimate are perfectly correlated with each other, but — as he understood intuitively — this is very unlikely. If we go to the other extreme, and make the assumption that all the estimates were independently derived, the overall error falls to 6.1 per cent.[15] In practice neither extreme is likely to be correct, and the true margin of error would lie somewhere between these limits. Without more information on the independence of the basic data we cannot say precisely where, but given the great variety of sources used by Kuznets, it is highly probable that it would have been closer to the lower than the upper bound, almost certainly below 10 per cent.

American reluctance to follow the treatment of errors by statisticians in the United Kingdom had not changed by 1955, when Raymond Goldsmith completed his mammoth study of savings in the United States. He observed with some passion that: 'the presentation of quantitative data without an indication of their probable or possible error is of limited value,' but accepted that this was what was normally done. He suggested that one of the reasons for 'this lamentable state of affairs' was 'the absence of any theory for the measurement of errors in quantitative economic data which are not derived from probability samples.'[16] Goldsmith cited the work of Wesley Mitchell *et al* in 1921 and of Kuznets in 1941 as the only attempts in the United States to assess the margin of error in economic series, but made no reference to any of the several examples by then available in Britain. He concluded the discussion of the errors in his own estimates with a simple guess as to the probable error in the overall total of saving, with no attempt to derive this from errors in the components by some appropriate statistical procedure.

After this brief interlude the topic lapsed once again and we are not aware of any later attempts to adopt or develop the work of the pioneers. A number of estimators and official agencies, including the Central Statistical Office in the United Kingdom, gave overall subjective reliability grades for major components and aggregates of the national income, but did not use any formal procedure for deriving these from the component errors.[17]

---

[15] Calculated by applying the subjective errors given by Kuznets, *National Income*, pp. 511–12 to his estimates of income originating, using the procedure described in section III below; see particularly fn. 34.

[16] Goldsmith, *Study of Savings*, II, p. 129.

[17] CSO, *Sources and Methods*, 39–42; Feinstein, *National Income*, pp. 20–22.

Modern investigators in the United States appear not even to have done this. When Goldsmith served as commentator at an NBER conference for a paper by Robert Gallman on the United States capital stock he chided him for not discussing: '… at least verbally, the probable margins of error in the main component series if he is not willing to follow Simon Kuznets's bold example of half a century ago … of indicating quantitatively the range of the margins.'[18] Instead, authors typically provide a general appraisal of the reliability of their estimates, but with no attempt to indicate the scale of the likely margins of error.[19]

If, despite this, we now wish to make a renewed attempt to promote the use of subjective assessments of margins of error, it will be necessary both to explain and clarify what this involves, and to evaluate the intrinsic merits of the procedure. We begin in Part II by briefly distinguishing the different types of error that can arise in quantification of historical series, and then discuss the virtues and limitations of two of the three separate but complementary techniques which can be used to assess their reliability.

In Part III we introduce a less familiar third approach: the assignment of subjective margins of error. We explain the statistical basis for this procedure, explore the underlying theoretical assumptions about the nature of the errors, and consider the means by which different elements contribute to the overall margin of error. In our view the fundamental procedure is sound, and if applied with due caution will yield a measure of error which provides useful additional information to both the compilers and the users of statistical series. Furthermore, the time taken to assign appropriate error classes to the component series can only be a small fraction of that required for the collection and processing of the data, and this final step should be regarded as an essential element in any publication of new historical estimates.

We also show how proper understanding of these points can help to achieve a higher degree of reliability in the process of estimation, whereas failure to appreciate them may lead to unnecessarily large errors. In conclusion, we revert in Part IV to the debate about the industrial revolution, and explore the implications of applying our method to Hoffmann's original estimate of the growth of

---

[18] Gallman, "United States Capital Stock," p. 210.

[19] This applies, for example, to estimates of output and productivity presented by Gallman, "Gross National Product;" Kendrick, *Productivity* Trends; Weiss, "US Labor Force Estimates," and Harley, "British Industrialization." It is equally true of British estimates; for example, those compiled by Deane and Cole, *British Economic Growth* and Crafts, *British Economic Growth.*

industrial production, around which so much of the controversy about growth during British industrialization has been centred.

II
Types of error

As a first step it will be helpful to consider three different sources of error: of measurement, of omission, and of procedure.[20]

Errors *of measurement* can arise in a variety of different ways. There may be errors in the basic source that the historian is using, either because of evasion or falsification on the part of the individuals and enterprises who provide the original data, or because of errors by those who collect and record them. For example, firms might understate sales or profits where these are to be used as a basis for taxation; or enumerators might omit or misclassify part-time female labourers in a nineteenth-century census of occupations. In other cases the basic data may come from samples and so be subject to sampling errors.

Even if the primary source is accurate, it may not be available in precisely the form required, so that some adjustment is necessary; the resulting figures are then likely to be less reliable than those taken directly from the primary source. The original data may be available only for selected years – as with a census – so that estimates for other years have to be derived by means of interpolation or extrapolation; or they might be incomplete, so that they have to be 'grossed-up' to cover missing sectors; or they might be given only as aggregates which have to be allocated in some way between two or more items.

Alternatively, the data source may be accurate and complete, but may not match precisely the concept or definition in which the historical investigator is interested. In that case, data originally designed and assembled for one purpose (company accounts, bank records, taxation, unemployment insurance) must be adapted to comply with different purposes and concepts, such as those of the national accounts. We may also include in this category measurement errors that are not inherent in the underlying data, but are introduced because the historian is guilty of mistakes, for example, entering the wrong numbers, adding them up

[20] See Morgenstern, *Accuracy of Economic Observations* and Goldsmith, *Study of Savings*, II, pp. 129–49 for more detailed discussion of these issues.

incorrectly, or failing to notice a change in definition recorded in an obscure footnote.

Secondly, there may be errors *of omission*, as when some components of a specified series are left out, either because the historian could find no source on which to base an estimate, or because they were thought too small to matter. Historical estimates of money earnings will typically be unable to obtain information for all wage earners; and indices of the cost of living of the working class will be unable to measure the changes in the retail prices of all relevant goods and services.

Thirdly, there are numerous ways in which the compiler of a series may introduce avoidable errors *of procedure*. These include adoption of an incorrect method of interpolation between benchmarks; the choice of an inappropriate base year for the construction of an index number; or the use of an unsuitable price index; for example, deflating an estimate of GDP by an index of wholesale prices of food and raw materials. Some components may be omitted in error, while others may be counted twice; or an inconsistent definition may be applied in estimating a particular component.

All these various types of error may in turn be either random or systematic. Random errors imply that the estimates may be either too low or too high, and such errors can generally be assumed to be scattered evenly above and below the true (unknown) value. These random errors are the central focus of our subsequent approach and will be discussed more fully below. By contrast, systematic errors occur when all the observations for a given series are too high, or too low. This might happen either because of consistent omission or double counting, or because of a persistent source of error in the basic data. Examples of systematic understatement include estimates of deaths derived from Church of England burial registers without allowance for non-conformists; or of household expenditure derived from self-reported budgets without correction for the typical omission of significant amounts spent on drink and tobacco. Systematic overstatement is illustrated by measurement of earnings without deduction for time lost through illness or unemployment, or of money supply statistics without correction for inter-bank deposits or coin held by the banks.

Whenever the investigator is aware of the existence of such systematic errors, the appropriate procedure is to correct the series as accurately as possible. This is particularly important where estimates are being made for a single date, or where absolute levels are important. It may matter less where the focus is on comparisons over time (or space), if it can be assumed that the relative under- or overstatement is broadly the same for all the relevant dates (regions).

Where corrections are required, they may have to be made on the basis of very limited or imperfect data but the essential principle to follow 'is that it is better to move, however uncertainly, in the right direction than to stand firmly in what one knows to be the wrong position.'[21] The extreme case of this arises where there is a complete lack of information on a missing component. Even here it is better to include a best guess than to ignore the item completely: 'there is no reason to regard zero as a closer approximation to the truth than a reasonable guess.'[22] Once such necessarily rough corrections are made, the known systematic errors will be effectively eliminated, leaving random errors which can then be handled by the procedure described in Part III.

Appraisal of reliability

There are three principal approaches to an assessment of the reliability of quantitative estimates: sensitivity analysis, comparison of alternative estimates, and assignment of subjective margins of error. Each has its advantages and can be applied independently. Where it is possible to use all three together they complement one another very effectively, and can collectively contribute to a more thorough appreciation of the reliability of the estimates.

Sensitivity analysis involves the construction and comparison of alternative variants in order to determine the extent to which the final result is dependent on particular assumptions or procedures. One illustration is the use of different budget weights for urban and rural areas in both the north and south of Britain to construct four separate cost of living indices over the course of the industrial revolution. This established that these different weights 'matter very little to the long-term trends in the cost of living.'[23] Another is seen in Thomas Weiss, who gives three variants of his conjectural estimates of the output of the American economy before the Civil War. Two of these rely on the same basic estimates of the labour force, but show the effect of incorporating different assumptions about the growth of output per worker.[24] The sensitivity approach can be extremely informative. Its limitations are that it can only explore the implications of alternative assumptions explicitly introduced by the investigator and it does

[21] Feinstein, *National Income*, p. 153.

[22] An observation by Austin Robinson in his foreword to Deane, *Colonial National Income*, p. ix. See also Goldsmith, *Study of Saving*, II, p. 141.

[23] Williamson, *British Capitalism,* pp. 210–13.

[24] Weiss, "US Labor Force Estimates."

not provide any way of combining the errors in individual components into a single comprehensive indicator of the overall scale of error.

The comparative procedure can be applied wherever it is possible to measure the same item in two genuinely independent ways. Perhaps the best-known example of this type of approach is the compilation of independent GDP estimates drawn from income, expenditure, and output data. The difference between such estimates – the residual error or statistical discrepancy – is a valuable indication of the reliability of the three series.[25] If all the estimates are assumed to be of equal reliability a compromise measure may be obtained by taking a simple average. However, if – as is more likely – some components are considered to have greater reliability than others, a more refined technique can be adopted, leading to an average in which the more reliable components are given greater weight. This was first proposed by Stone *et al,* and has recently been developed under the name of 'balanced' estimates by Martin Weale and others.[26]

The comparative approach can also be fruitfully adopted in relation to a variety of other estimates. Independent series for capital formation can be derived from data on production and international trade, and from financial information in company accounts.[27] In the sphere of foreign trade, the reliability of the statistics can be evaluated by comparing exports recorded by one country with the import statistics of its trading partners. The results reveal marked discrepancies when comparison is made at the level of individual countries; but the data are more accurate when the comparison is made on the basis of the total value of a country's trade.[28] On the capital side, comparison of estimates of interwar capital flows, on the basis of balance of payments statistics for all the leading countries, showed that the capital movements recorded by the creditor nations were consistently lower than those of the debtor nations. In this case, the substantial discrepancies were judged to be too great to be explained solely by errors in the data, and were interpreted as indicating large-scale evasion of the exchange controls in ways which distorted the current account data.[29]

---

[25] US Department of Commerce, *National Income Supplement*, pp. 66–7; Feinstein, *National Income*, pp. 12–20; Britton and Savage, "Three Measures."

[26] Stone, et al "Precision;" Weale, "Reconciliation of Values;" Solomou and Weale, "Balanced Estimates;" Greasley and Oxley, "Balanced versus Compromise Estimates;" Sefton and Weale, *Reconciliation of National Income*.

[27] Kuznets, *Capital*, pp. 602–12; Feinstein, *Domestic Capital Formation*, pp. 230–36.

[28] Morgernstern, *Accuracy of Economic Observations,* pp. 163–80; Federico and Tena, "Accuracy of Foreign Trade Statistics."

[29] Feinstein and Watson, "Private International Capital Flows."

A variant of the comparative approach is sometimes used in connection with the national accounts and other official statistics where it is possible to compare successive estimates of the same item.[30] However, the later estimates are usually based on more complete and reliable data, so that the discrepancy between these and earlier estimates may not indicate much about the reliability of the final version. More generally, it is sometimes possible to compare alternative estimates of a given series by different authors. This is also a potentially useful guide to the reliability of the series, particularly where the compilers have worked with different sources, though it must be done carefully to allow for any discrepancies in definition or procedure.

Where these various forms of comparison can be applied, they provide a useful indication of the overall reliability of the estimates, reflecting the net extent of errors of all types in both data and construction. Unfortunately, however, it is often not possible in historical research to find more than one usable source of data for any given series, and the opportunities for genuinely independent comparison are thus limited. Moreover, even where comparison is possible, it cannot be assumed that the two series necessarily set upper and lower bounds to the true value; it is possible that both estimates could be too low or too high.

III
Subjective margins of error

The third approach, and the one with which we shall be principally concerned, requires the compiler of an estimate to attach margins of error to each of its components, and then to derive an overall margin of error for the total or series. These estimates are not normally derived from random samples, so it is not possible to calculate objective confidence intervals by means of standard statistical procedures, and the margins of error must necessarily be rough and subjective. The strongest explicit case against this procedure we have found was made in 1953 by Milton Gilbert, one of the leading figures in the development of the national product estimates of the United States. His initial objection was: ‘that we simply do not know the size of the margins of error in the estimates with enough accuracy to quantify our judgments. The reason is that in the complex of factors that might lead to inaccuracy of the statistics, there are no measures of the errors

[30] US Department of Commerce, *National Income Supplement*; Cole, *Errors in Provisional Estimates*.

arising out of most of them, and hence no way to assign them weights so as to arrive at a combined margin of error.'[31]

However, these and other objections were made by Gilbert in the specific context of the contemporary estimates of US national income compiled by a large official agency from reported benchmarks. In our view, they would have much less force in relation to historical estimates typically compiled by a single investigator using less complete and reliable sources. The individual who has compiled such a set of estimates is far better placed than anyone else to evaluate them. She will be most familiar with the quality and completeness of the sources from which the estimates were derived, and of any alternative information that was available; and will know in detail the nature and reliability of the various adjustments she has made to the original data. She will also have most insight into any checks that were undertaken at various stages in the construction of the estimates, and most knowledge of any comparisons of her results with independent estimates. On the basis of all this information she can judge the 'range of reasonable doubt attaching to the estimates.'[32] If this is done for each separate component (for example, the wage in each sector), the results can then be combined to form an overall judgement of the reliability of the larger aggregates.

However problematical such subjective assessments of unknown errors may be, they are much more informative than general statements formed from some favoured permutation of stock phrases (these estimates are very: 'approximate', 'imperfect', 'unreliable', 'tentative', 'uncertain', 'fragile'; they are: 'a best guess', 'a rough guide', 'an order of magnitude', 'a crude indication'; or, very occasionally, they are: 'reasonably reliable', 'broadly acceptable'; and so on). Potential users are surely much better served if they are told, for instance, that the margin of error assigned to an estimate of 8 shillings for average weekly cash wages for farm workers in England in 1770 is ± 25%. Such an assessment can be evaluated by users of the data in a way that is not possible with qualitative statements; for example, by considering the frequency of references to wages of less than 6 shillings or more than 10 shillings. If the user finds evidence to suggest that such values are actually quite common she is immediately able to revise the proposed margin of error.

To understand the basic principles of the procedure we recommend for dealing with random errors, imagine a hypothetical situation in which the investiga-

[31] Gilbert, *Studies in Income and Wealth*, III, pp. 6–8.

[32] CSO, *Sources and Methods*, p. 40.

tor is able to observe not just a single estimate of the average wage in the building industry in say 1930, but a number of independent estimates. This might be done, for example, by sending out a team of investigators, each of whom was required to make her own estimate by enquiries in a random sample of towns or firms all covered by a national wage agreement. For each such estimate there will be a corresponding error, and it is thus possible to build up a pattern of these estimates and their associated errors. Imagine that the true (unknown) value of the average annual wage in the United Kingdom building industry in 1930 was £150. Some of the independent estimates will be less than £150, some will be more; either because of errors in the data recorded by the investigators or their informants, or because of specific circumstances in the individual labour markets which cause the local wage to deviate from the average set by the national agreement.

The crucial feature of the law of errors on which the Stone-Desai-Chapman procedure relies is that the distribution of the estimates will be distributed in a roughly symmetrical pattern either side of the true mean of £150. It then follows that the *errors* in the estimates of the component parts will also be roughly symmetrical, and the mean error (i.e. the difference between the several recorded figures and the true figure) will be zero.

The measure of the dispersion of such hypothetical errors around the mean national wage is known as the standard error. In accordance with classical statistical theory, if the errors are distributed in this way, a distance of one standard error to either side of the mean will include roughly two-thirds of all observations. Similarly, a distance of two standard errors will include roughly 95 per cent of the observations, with 2.5 per cent in each of the tails to the left and the right of the distribution. In more formal terms, the assumption is that the distribution of the errors is normal with mean zero, but all that is required to apply the method is that it should be roughly normal, i.e. it should not be markedly skewed.[33]

The relationship between these standard errors and the subjective margins of error assigned by the estimator thus depends on the degree of confidence attached to the estimator's judgement about the margin of error. They might, for example, be assigned on the basis that the chances are roughly one in three that the true value lies outside the specified range of error. This is equivalent to saying that the margin of error above or below the mean corresponds to one stan-

[33] If there was reason to believe that in a particular case the distribution was markedly skewed (for example, for a very small item for which the estimates could not be less than zero) this could be dealt with by taking the logarithm of the estimates; see Stone, "Two Studies," p. 69.

dard error. If a higher degree of confidence is required (at the cost of a wider range), the margins of error can be assigned on the basis that there is only one chance in twenty that the true value lies outside the specified limits. This represents a 95 per cent probability level, and implies that the margin of error on either side of the mean is analogous to two standard errors.[34]

We have already referred to the work of Chapman, and her study of wages and salaries in the United Kingdom in the interwar period can be taken as an excellent illustration of the recommended procedure. She assigned one of the following reliability grades (or error classes) to each of her component estimates of (a) the numbers employed in each sector and of (b) their average earnings:[35]

Reliability grade	*Margin of error*	*Average margin of error*	*Standard error*
(1)	(2)	(3)	(4)
A. Firm figures	± less than 5%	± 2.5 %	± 1.25 %
B. Good estimates	± 5% to 10%	± 7.5 %	± 3.75 %
C. Rough estimates	± 10% to 25%	± 17.5 %	± 8.75 %
D. Conjectures	± more than 25%	± 40.0 %	± 20.00 %

To simplify the calculations, Chapman worked with the average margin of error in each of the four reliability grades (assuming that the average for Class D was 40%). Her judgements were made with 95% probability, so each of these average margins of error is thus the equivalent of two standard errors (see columns 3 and 4 of the above table).

The next step is to combine the standard errors for the components of a series in order to derive an assessment of the overall error in the estimate. Given the symmetric nature of the error distribution, some of the component estimates will be above the true (but unknown) mean for the sector, others will be below. It follows, therefore, that if the component estimates are independently derived

[34] In relation to the work of earlier investigators cited in Part I, the margins of error assigned by Mitchell, et al, *Income in the United States*, corresponded to only two-thirds of a standard error, with a 50% probability level; whereas those adopted by Deane, *Colonial National Income*, corresponded to three standard errors at the 99% level. Kuznets adopted a somewhat different approach, stating that in assigning maximum errors his team were concerned with how large the error could be, but without indicating the degree of probability with which these judgements were made; see Kuznets, *National Income*, II, p. 503, and the criticism of this in Stone, "Two Studies" p. 68. This means that the approach outlined above cannot be directly applied to his estimated errors, though it might perhaps be assumed that he intended a very high degree of confidence so that the margins could be regarded as analogous to roughly three standard errors either side of the mean.

[35] Chapman, *Wages and Salaries*, p. 231.

the errors will to some extent offset each other. The proportionate error in the total will thus be smaller than the weighted average of the proportionate errors in the components. The precise extent of the compensation will depend on factors such as the number of individual sectors, the relative size of each sector, and the margin of error in each of the component estimates. The greater the degree of interdependence, the smaller will be the compensating effect of aggregating the component series. Provided that the errors are less than perfectly correlated the error of the whole will always be less than the error of the parts.

The formula for combining the standard error of the whole from the standard error of the parts is derived in the Appendix, using both probability theory and simple differential calculus. The basic formula (from equation A7) is:

$$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + 2r_{xy}\sigma_x\sigma_y} \qquad (1)$$

which states that the standard error in V is the square root of the sum of the variances of X and Y plus the error derived from the interdependence of X and Y. The extent of the interdependence is indicated by r_{xy}, the correlation coefficient of the two components. If the two components are completely independent, the correlation coefficient will be zero; if they are consistently interrelated, the correlation coefficient will be either –1 (if an error in one item is offset by the error in the other) or +1 (if the error in one term is compounded by an error in the other); if the errors are related, but not in a monotonic fashion, the correlation coefficient will be non-zero, but will fall between the two extremes.

To illustrate the procedure, let us take the simplest case of an estimate that is the sum of two components; for example, national income derived as the sum of wages and profits. Let the estimate of wages and profits be 800 and 200 respectively, and their margins of error, 20% and 40% (assessed at 95% confidence intervals). The absolute standard errors are thus 80 and 40 respectively. There are then four primary possibilities. If the errors in wages and profits are independent, so that the correlation between them has a coefficient of zero, the standard error of the whole equals $\sqrt{(80^2 + 40^2)} = 89.44$. If the errors are perfectly positively correlated, such that the correlation coefficient is 1, the standard error equals 80 + 40 = 120; this is identical to the result obtained by taking a simple weighted average of the proportionate margins of error. If the errors are perfectly negatively correlated, so that the correlation coefficient is –1, the standard error of the whole equals 80 – 40 = 40. Finally, if they are partially correlated, with a correlation coefficient of, say, 0.5, it equals $\sqrt{(80^2 + 40^2 + (80 \times 40))}$ = 105.83.[36]

Since the initial margins of error were assigned at the 95% level, the absolute margin of error in the whole is equal to two estimated standard errors. Thus, the proportionate margins of error range between 8.0% $\left(\frac{2 \times 40}{1000} \times 100\right)$, when the errors are interdependent and fully offsetting (r = –1), and 24.0%, when they are cumulative (r = +1), compared to 17.89% when they are independent (r = 0).

In order to investigate systematically the full range of possibilities it is necessary to adopt a slightly more formal approach, and this is set out in the Appendix. The discussion covers both the relationship between the total and the components, and the effect on the results if the errors in the components are either independent of each other or, alternatively, positively or negatively correlated with each other. The Appendix begins with the simplest case where the total is equal to the sum of two or more components. It then proceeds to cover more complex cases including products and quotients, weighted averages, and ratios of weighted averages.

Readers who wish to skip this can proceed directly to Part IV, where the procedure is applied to a recent estimate of the rate of growth of British industry during the industrial revolution.

IV

Errors in the measurement of British industrial production, 1780–1831

As we have noted in the introduction, the rate of growth of industrial production during the British industrial revolution has been the subject of sustained controversy. It thus represents a suitable test case for the procedure that we are recommending for the assessment of margins of error. Our aim in this final section is to illustrate the procedure by applying it to the original index of industrial production compiled by Hoffmann in the 1930s and published in English in 1955. It is this index that has been challenged in the subsequent revisions by

---

[36] Clearly, there will be as many possible outcomes as there are possible values of r_{xy}. However, they will all fall within the range provided by $r = \pm 1$ and, in the case of indefinite interdependence, small perturbations around 0.5 will generate only small differences in the calculated margin of error.

Harley, Crafts, Jackson, and others.[37] It is pertinent, therefore, to ascertain whether or not these revisions lie within the bounds of the probable margins of error in the Hoffmann estimates. We make this assessment for the successive periods 1780–1801 and 1801–1831 that are at the heart of the debate about the pace of economic growth, and restrict our attention to the industrial index, excluding building.

Our concern is solely with the margins of error in the index as it was constructed by Hoffmann, i.e. with the effect of the probable errors in the indicators and the weights used for the index. There are other important issues that must also be considered in any evaluation of an index, notably the choice of base year and whether a Laspeyres or Paasche formula (or some variant such as the Fisher ideal index) is used for weighting the indicators. However, these are best handled by other procedures; for example, by constructing alternative indices with different base years or different formulae, and comparing the outcomes to determine the sensitivity of the results to these alternatives.

Before proceeding to the calculation of the margins of error it is worth noting a few points regarding the methods used by Hoffmann to construct his index. It is a weighted arithmetic mean of physical indicators of industrial production, so the basic formula for any given year 1, where year 0 is the base year, is thus:

$$\frac{\sum_{i=1}^{m}\left(\frac{Q_{i1}}{Q_{i0}}\right)P_{i0}Q_{i0}}{\sum_{i=1}^{n}P_{i0}Q_{i0}} \times 100 \qquad (2)$$

where Q_{i1} / Q_{i0} is the level of the output indicator for industry i in year 1 relative to the level in the base year, and this relative is weighted by the industry's share in value added in the base year. There are a total of n industries, but only m of these have independent quantity relatives.

Hoffmann initially constructed indices for six sub-periods from 1700 to 1913, with a separate set of weights for each of these sub-periods. For the dates in which we are interested, 1780 falls within the period 1760–1800 for which the available indicators are combined with 1783 weights; 1801 within the period 1800–30 for which the base year is 1812; and 1831 within the period 1831–60 for which the base year is 1850. Hoffmann linked the indices for these and other

[37] Harley, "British Industrialization;" Crafts, *British Economic Growth*; Crafts, Leybourne and Mills, "Trends and Cycles;" Crafts and Harley, "Output Growth;" Jackson, "Rates of Industrial Growth."

sub-periods by extending the weighted average for each sub-period by one year to form an overlapping year with the following sub-period, and then used these overlapping years to form a single index with 1913 = 100.[38]

In order to assess the margins of error in Hoffmann's index, we are sometimes forced to make certain assumptions regarding his methods where he is silent on the precise procedure he adopted. This, of course, is not a problem that would be encountered by an author assessing the margins of error in her own estimates. Hoffmann provides a reasonably full account of the sources for his quantity indicators, but he gives very little information about the sources and methods he adopted to obtain his base year weights. However, it seems clear they are intended to be measures of net output (value added) based on two sources: estimates of industrial output and of wages and salaries.[39] A crucial feature of his method involves those industries for which no indicators are available. As indicated by his weights, the covered industries accounted for 56.4 per cent of total industrial production in 1783, 67.5 per cent in 1812, and 70 per cent in 1850, but the proportions are lower than this for years in the earlier part of each sub-period for which the indicators are not yet available. Thus in 1780 the known indicators account for only 43.3 per cent of total industrial production (excluding building). Hoffmann allocated the weights for the missing sectors pro rata to those covered; for example, the weight initially assigned to coal in 1783 was 6 per cent, but in constructing the weighted average for 1780 this was raised by 130.9 per cent (100 / 43.3) to 13.9 per cent, and similarly for the remaining covered sectors. This effectively assumes that the growth rate of the unknown industries was the same as that of those covered, and the appropriateness of this (implicit) assumption is one of the key issues in the subsequent debate.

We shall approach the exercise in two stages. In the first stage we calculate the error in the level of the index at the beginning and end of the period over which we wish to measure the rate of growth. For the first of our two periods we thus have 1780 combined with 1783 weights and 1801 with 1812 weights. For each year this requires us to combine the errors in the quantity relative for each

---

[38] Hoffmann, *British Industry,* pp. 15–19. This procedure requires a value for each indicator in the base year, and Hoffmann does not state explicitly how he dealt with those indicators which start at a later date; for example, in the period 1761–1800 the base year is 1783 but the indicators for silk yarn and cloth, beer, and tobacco are available only from 1787. However, it seems consistent with his general procedure to assume that Hoffmann extrapolated back to the base year for these missing indicators on the basis of the weighted average of the known indicators.

[39] Ibid., pp. 17–23.

industry at that date (Q_{i1} / Q_{i0}), and the errors in the value added weight for each industry ($P_{i0}Q_{i0}$ / ΣP_0Q_0). As the weights for the beginning and end dates are not the same, the former corresponds to the terms X_i and Z_i in equation (A32) in the Appendix, the latter to the terms w_i and y_i. In the second stage we calculate the total error involved in taking the ratio of the weighted index at the second date to the ratio at the first date:

$$\frac{\sum w_i X_i}{\sum y_i Z_i} \tag{3}$$

Exactly the same procedure is followed for the second period, 1801–31 where 1801 is combined with 1812 weights and 1831 with 1850 weights.

Two aspects of the potential errors in the estimate have to be considered in each stage: first, the outcome assuming all the errors in the components are completely independent; second, the corrections required if there is any form of relationship between the errors in the components. The overall error in the ratio will be comprised of the error in each of the weights and the error in each of the indicators. It can easily be shown that the total variance is additive in the variances of the four components of the ratio; the derivation of the appropriate formula and its particular structure are set out in detail in the Appendix.

In order to make the error formulation operational, it is necessary to attach margins of error to each of the components of the ratio in equation (3). We will now consider each in turn, starting with the errors, X_i and Z_i, in the quantity relatives. The crucial point here is precisely the fact that these are relatives. The logic of the index number formula in equation (2) requires us to assess the margin of error in the level of the quantity indicators for each industry in the given year *relative to its level in the base year*.

This has the important consequence that we are concerned only with the random or stochastic errors in the data; for example, those created by mis-recording of the numbers by clerks in the Board of Customs and Excise, or those in some author's 'best guess' as to the level of pig iron produced in a single year. We can reasonably assume that any systematic sources of error in the level of the indicator — such as those arising from the deliberate concealment of output by soap and other manufacturers to evade Excise Duty, or the use of data for a single region (such as the woollen cloth produced in the West Riding of Yorkshire) to represent national output — will have the same proportionate effect on the given year and on the base year, and can thus be ignored.[40]

[40] If this is not the case, then the procedure advocated in Part II above would require that we should attempt as best we can to eliminate any identified systematic error. If, for example, we

The proportionate margins of error assigned to each of the quantity relatives for the three years 1780, 1801 and 1831 are shown in Table 1. Each error is to be interpreted using the rule of two standard deviations, namely that there is a 95 per cent probability that a range of two standard deviations around the measured relative will contain the true (but unknown) relative.

We turn next to the errors in the weights, w_i and y_i, i.e. in the relative size of each industry in the relevant base year. We assume that Hoffmann made estimates of the value added in each sector (including those which he was unable to incorporate in his index because no quantity indicators were available) and summed these to obtain the total for industry as a whole. In this case, the weight for industry *i* at date 0 will be equal to the ratio between the value added in that sector and the sum of the value added in all sectors, in the form:

$$\frac{P_{i0}Q_{i0}}{P_{i0}Q_{i0}+\sum_{j\neq i}^{n}P_{j0}Q_{j0}} \qquad (4)$$

Our task is thus to assign a margin of error to the weight for each industry that combines the errors in the value added of the given industry and all other industries. The error in each weight in equation (3) will combine the error in the value added in that sector and the error in all other sectors (which is itself the combined value of the errors in each of the other sectors).

We then have to deal with the problem created by the industries for which no quantity indicators are available at the given date. As noted above Hoffman implicitly imputed the weights for these missing industries to those that he could cover. There are, therefore, three possible ways in which the associated margins of error might be handled. The weights for each of the known indicators could be grossed-up by the appropriate proportion (as illustrated above for coal); all of the missing industries could be treated as a single sector, with its own weight, and an indicator equal to the weighted average of the known industries; and each of the missing industries could be treated individually, each with its own weight, though all with the same indicator based on the known industries.

---

have reason to believe that evasion of excise duty was diminishing over time because of more effective inspection, then an indicator based on the duty would overstate the growth of the industry and we should make an appropriate correction for this. Note, however, that the attempt to do this is liable to generate a higher level of stochastic error. For example, if the level of soap output in 1801 relative to 1783 is raised by 10 per cent to allow for the reduction in evasion, this correction is itself measured with a margin of error — the true correction could range from, say, 5–20 per cent — and the error margin around the corrected figure should be increased accordingly.

None of these are entirely satisfactory. The disadvantage of the first method is that it complicates the determination of the margin of error of each of the weights and the indicators for the known industries, since the error in the former must also cover the error arising from the grossing-up process, and the error in the latter must cover its imputation to other industries. The weakness of the second method is the sheer scale of the single sector that it creates: in 1783, for example, the sum of the missing industries has a weight of 56.7 per cent. Since the overall procedure is sensitive to the size of the weights, a single sector of this magnitude would severely distort the calculation of the overall error.

The drawback to the third procedure is that it requires individual weights for each of the missing industries, but if this problem can be overcome this is the most appropriate option in the present context of an index number. Some, but not all, of these missing weights are given by Hoffmann for a later date, and those missing for earlier years can be derived on a proportionate basis from his estimates for the nearest subsequent year. This leaves a substantially smaller residual covering clothing, brick-making, glass, pottery, and a few other small industries that are omitted from the Hoffmann index throughout. We adopted this third procedure and the final margins of error derived on this basis for the weights in the three base years are shown in Table 1. Each of these errors is again to be interpreted using the rule of two standard deviations.

It will be evident from Table 1 that the margins of error attached to the weights are larger than those for the indicators, and that the margins of error attached to each sector's value added was set lower for 1850 than for the earlier years. The margins of error on the indicators range from ±10% for those sectors (such as cotton, beer, tobacco) for which either annual customs or excise data suggest a reasonably firm basis for year-to-year comparisons. For sectors in which Hoffmann's data were generated from reasonably good private sources (e.g. iron and steel after 1800; wool and linens) we assigned margins of ±25%. Finally, for those sectors for which the information is less than firm, or for which lack of information caused us to assign the industry-wide growth rate, we have assigned margins of error of ±50%.

We applied only three margins of error for Hoffmann's value added estimates: 25 per cent; 50 per cent; 75 per cent. In the earlier years, we judged that the quality of Hoffmann's estimates were seriously restricted by the quality of the information available — we therefore set most value added estimates at ±50%; for those sectors for which the data appeared to us even more fragile, the margin of error was set at ±75%. For 1850, we lowered the margin of error for most sectoral value added estimates to ±25%, although we maintained the high relative error levels for the weaker sectors, setting it at ±50%.

The higher error margins assigned to the weights reflects our judgement that the statistical base for measuring value added is simply less reliable than for levels of physical output. Partly this is due to the recognition that the indicators are measuring errors relative to each other, rather than in absolute terms, as with the weights. Similarly, our decision to reduce the errors attached to the value added measures for 1850 reflects a judgement that the statistical sources used by Hoffmann to establish his weights were more reliable for the later date. This improvement in collection methods and reporting is most important in measuring the level of value added in each sector. It makes less of a difference in the indicator series, which are almost all derived from official statistics, such as Excise returns and import figures, which are considered to have much the same level of stochastic error in each period. For this reason, we have taken the conservative approach of not generally reducing the errors of the indicators for 1801–31 relative to 1781–1801. However, for those sectors for which improved information became available between 1780 and 1801 (or for which it was possible to use sector-specific rather than industry-wide data), we have lowered the attributed margins of error.

Given these margins of error, we can calculate the overall margin of error under the assumption that each entry in the ratio is independent. The formula (reproduced from equation A33 in the Appendix) is:

$$\sigma_F^2 = \Sigma(F_w\sigma_{wi})^2 + (F_X\sigma_{Xi})^2 + (F_y\sigma_{yi})^2 + (F_Z\sigma_{Zi})^2 \qquad (5)$$

where F_w symbolises the impact of an error in the weight, w_i, on the index, F.

This can be re-written to give the proportionate error in the ratio (see equation A34 in the Appendix) as:

$$\left(\frac{\sigma_F}{F}\right)^2 = \Sigma\left(\frac{X_i}{\Sigma w_i X_i}\sigma_{wi}\right)^2 + \Sigma\left(\frac{w_i}{\Sigma w_i X_i}\sigma_{Xi}\right)^2 + \Sigma\left(\frac{Z_i}{\Sigma y_i Z_i + Z_i}\sigma_{yi}\right)^2 + \Sigma\left(\frac{y_i}{\Sigma y_i Z_i + y_i}\sigma_{Zi}\right)^2 \qquad (6)$$

This looks formidable but is relatively simple to calculate once a suitable spreadsheet has been established and the detailed margins of error assigned as in Table 1.

As the next step in the calculation we must drop the assumption that each component is independent, and correct for the effect of any possible relationships between parts of the index. There are, in principle, ten pair-wise combinations in the system as a whole. Of these, two arise between the quantity indicators and two between the weights *within* either the numerator or the denominator (r_{xx}, r_{zz}, r_{ww}, and r_{yy}). One arises between the weights and one between the indicators *between* the numerator and the denominator (r_{wy} and r_{xz}). The final four arise between the weights and the indicators *between* the numerator and the denominator (r_{wx}, r_{yz}, r_{wz}, and r_{xy}).

It is only the first four of these with which we need to be concerned in the present exercise. The fifth possible relationship can be ignored because there is no reason to believe that any error in the estimated share of any industry in one base year will be related to a corresponding error in the share of that industry in the following base year. The sixth possible relationship need not concern us because — as already argued — we do not think that any error in the relative quantity indicator for one date will be systematically related to the error in that indicator for the second date. Finally, the last four possible forms of interdependence can be ignored because Hoffmann's weights are estimated independently for each date; he did not rely on quantity (and price) indicators to extrapolate the value added from a single benchmark to other base years.

The first two of the four relevant interdependences relate to interdependence between the quantity indicators (r_{xx} and r_{zz}). This occurs quite frequently when the same indicator is used for two industries; for example, imports of raw cotton are used to measure the output of cotton yarn, and that series is then used (with an adjustment for net exports of yarn) to measure the output of cotton cloth. There are similar relationships in a number of other cases, including copper and copper goods, iron and steel and iron and steel products, woollen yarn and woollen cloth (after 1820), silk yarn and silk cloth, flour and bread, and leather and leather goods. In each such case we have assumed a perfect positive correlation ($r = + 1$) between the pairs of indicators, and the interdependence is measured, following equation (1) above, using the following formula:

$$2\, r_{xx} \sigma_{xx} F_x F_x \qquad (7)$$

and similarly for r_{zz} (see equation A23 in the Appendix).

The two remaining relationships relate to interdependencies between the weights (r_{ww} and r_{yy}). This occurs in the specific case of the Hoffmann index in the weights for the textile industries, where it seems probable from the sources available to him that he first made an estimate of the overall weight for the industry (cotton or wool or silk) and then subdivided this between the yarn and cloth components. If this is correct it would mean that there was always a perfect negative correlation between the two components.

Separate from these forms of correlation between the errors in the components of the index, there is a general accounting constraint created by the fundamental nature of the weights as proportions, and this has a further impact on the measurement of the overall error. Even though the weights begin as estimates of the absolute level of value added in each industry, they must be converted to a proportionate basis in order to be used as weights. It follows, therefore, that if the weight on one sector is too high, there must be compensating changes in the weights for each of the other sectors. In our analysis, we have assumed that any compensating changes will be shared proportionately by the remaining sectors. In order to make this further correction it is necessary to modify the formula given in equation (7) above. The corrected formulation is given in equation A35 in the Appendix.

Table 2 sets out the results of our analysis for the error in the rates of growth of industrial production over the two sub-periods, 1780–1801 and 1801–1831. We have shown the separate elements that make up the overall estimate of the variance in each of the two ratios. The standard error is then obtained as the square root of the variance, and the range around the original ratio (of 1801 to 1780 or 1831 over 1801) can be calculated by taking the original ratio plus or minus two standard errors. The corresponding annual percentage growth rates can then be calculated from these lower and upper bounds of the ratios.[41] On the basis of the rule of two standard deviations there is a 95 per cent probability that the expected true figure for either the ratio or the growth rates will lie within the corresponding ranges.

The results for the two sub-periods indicate that the margin of error for the Hoffmann index was around 12.5 per cent for 1780–1801 and 16 per cent for 1801–1831. It may be surprising that the error is larger for the later period, despite the fact that the margins of error in the indicators are identical, while the later period incorporates a lower weight (for 1850). Close examination of the component elements of the error structure reveal that the cause of this higher overall margin lies in the significantly larger size of the residual sector (clothing, bricks, and so on) in 1850 than at any of the earlier benchmarks (the size of this 'other' sector rose from 16.1% in 1813 to 22.9% in 1850). This finding emphasizes the sensitivity of the procedure to the existence of large sectors. It also suggests that a significant improvement in the certainty of the overall index could be made by improving our knowledge of the components and behavior of

[41] It will be noted that the calculated margin of error, although distributed symmetrically around the mean estimate of the ratio of output in the two years, does not produce a symmetric range around the calculated growth rate. This reflects the impact of compounding, whereby small changes in growth rates can create large changes in levels after 20 or 30 years. The range from the original ratio to the upper bound of the growth rates is, therefore, always smaller than the range to the lower bound.

this residual sector, both by reducing its size and reducing the error attached to the indicator.[42]

The decomposition of the overall error into its component parts suggests other general conclusions. The issue of interdependencies is clearly second-order (it changes the margin of error by less than 2 per cent in either sub-period), while the impact of incorporating the correction for compensating weights is substantial (reducing the size of the overall error by more than 10 per cent in each sub-period). Indeed, the errors attached to the weights in each index decline dramatically once the accounting constraint that the share of all sectors in aggregate output must sum to one is applied. This is entirely appropriate, since it means that any error in a weight will have a compensating response in all the other weights. As long as growth in the economy was relatively balanced across sectors, the shift in weights from one sector to another would have little impact on the overall measure of growth.[43] Thus, while the error in the weights was always set greater than the error in the indicators in our procedure, the net effect makes it clear that it is the error in the indicators that is crucial to the overall evaluation of error in the growth rates.

The net result of the calculation suggests similar margins of error around the growth rates derived from the original Hoffmann index for both sub-periods. The range in growth rates is 3.2 to 4.4 per cent p.a. for the first period; 2.2 to 3.3 per cent p.a. for the second. If our purpose were to determine whether there was a change in the rate of growth between the two sub-periods, the confidence intervals would suggest that the null hypothesis could be rejected at the 95 per cent level. However, the range in growth before 1801 is narrow enough to exclude the lower figures embraced by Crafts, Harley, Jackson, all of which hover around 1.9 per cent p.a. Whether or not the difference in the growth rates in Hoffmann and Crafts-Harley-Jackson is statistically significant will, of course, depend also on the size of the confidence interval around the latter estimate. A brief back-of-the-envelope calculation, assuming similar error margins for the

[42] The largest single sector in this residual group is clothing, which clearly grew more rapidly than industrial output as a whole. This industry is excluded from the revisionist indices of Crafts, Harley and Jackson, suggesting that their growth rates may be too low, especially in the 1801–1831 sub-period.

[43] Indeed, to emphasize this point, the one sector that contributed significantly to the overall error through the possible error in its weights was cotton, the fastest growing sector in the economy.

component elements of the Crafts-Harley-Jackson series as for the exercise shown here, suggests that the two confidence intervals would not overlap.[44]

Of course, any such comparison assumes that the margins of error assigned in Table 1 are appropriate. One of the singular advantages of this procedure is that it enables us to work out what the range of growth rates would be like if different margins of error were assigned. Thus, for example, it might be thought that our views on the quality of the data were too sanguine, and that all the assigned margins should be raised by 50 per cent (thus retaining the relativities in the errors both across indicators and weights).[45] The net effect of this change is to raise the overall margin of error of the weighted index from 12.5 per cent to 19.9 per cent for 1780–1801 and from 16.4 per cent to 25.6 per cent for 1801–1831. The range in the growth rates is also increased (to 2.7 to 4.7 per cent p.a. for 1780–1801; and 1.8 to 3.6 per cent p.a. for 1801–1831), thus reducing the statistical significance of any difference in the growth rates across the two sub-periods.[46] The mechanism by which these higher component margins translate into a larger margin of error of the index as a whole is through the larger errors assigned to the indicators; once again, the procedure of compensating weights reduces the impact of errors in the value added at each benchmark to comparative insignificance, while the effect of the interdependencies is again small.

We believe that these larger margins of error are too high for the Hoffmann index. However, a great advantage of the procedure that we are advocating is that it is no longer necessary to accept the limited range of options chosen by the researcher in his sensitivity analysis; if a critic prefers another structure of errors (which could involve changing relative errors within the system, as well as or instead of making scalar changes across all observations), the impact on the overall result can be found. The critic, in turn, needs to take the challenge of

[44] In this very simple exercise, which makes no pretence to being a full exploration of the differences between Hoffmann and the revisionists, we substituted the growth rate at the economy-wide rate excluding cotton and iron and steel for the sectors that lack indicators (thus reflecting one of the primary aspects of Harley's original critique); we retained the margins of error attached to both weights and indicators. The calculated growth rate for 1780–1801 on this basis has a range of 1.2% to 2.5% p.a., which certainly falls outside the range of the Hoffmann index for the same period. The margin of error in this revised index is 13.4%, slightly higher than for the fully developed Hoffmann model.

[45] Subject to the limitation that no error can be larger than 100%.

[46] Furthermore, increasing the component margins of error by 50% all around would cause us to revise the conclusion that there is a statistically significant difference between the growth rate of the Hoffmann index and the Harley-Hoffmann index (incorporating only the changes in the growth rates of the residual sectors).

assigning margins seriously and to allocate figures that are persuasive to other readers. This need for accountability is a great strength of the procedure for which we are making our plea.

APPENDIX

The basic procedure

This section derives the basic rules of procedure for estimating the margin of error for quantitative estimates. The rules are derived from mathematical statistics. They are relatively complex in their construction, yet have intuitive meaning. We start with a statistical estimate that is subject to random errors. Let us call the estimate N (for, say, national income) and the error, V. Since V is a random variable, it follows a particular probabilistic distribution, the exact shape of which depends on the assumptions made about the error structure of the estimate. It is standard to assume that errors are symmetric, continuous and centred on zero; i.e. that the estimate is a best guess but is probably in error by $\pm$ x per cent, where x can take any value between $-\infty$ and $+\infty$. Under these conditions, the distribution of errors will be normal, or Gaussian.[47]

Let us formally define the parameters of the distribution of the variable, V, as follows. The mean, defined as the expected value (or mathematical expectation) of the distribution produced by repeated samplings, is represented by E(V), or μ_v. In this particular case, as noted above, μ_v is assumed to be zero. The variance of V, represented by σ_v^2, measures its dispersion around this mean value, and is defined as the squared difference between the value of V for any given observation and the mean value of V for the population as a whole, again measured over repeated samplings of the distribution. Thus, for any random variable, V, with mean, μ_v, the variance σ_v^2 is given by:

(A1) $$\sigma_v^2 = E[V - \mu_v]^2,$$

where E[.] symbolizes the expected form. Taking the square root of the variance produces the standard error of the estimate, σ_v, such that the estimate may be written, N $\pm$ σ_v.

Assume now that our estimate of national income originated in two sub-estimates (wages and profits), each of which have been measured as accurately

---

[47] What follows does not depend on this particular distributional form; it applies to any probability distribution based on random errors (including the binomial, multinomial, Poisson, etc.)

as possible, but are nonetheless subject to random errors, designated as X and Y, respectively. The total error, V, is thus a linear combination of other random variables, X and Y, with population means, μ_x and μ_y, and variances, σ_x^2 and σ_y^2. The issue is how the particular distribution of the errors in these component parts shapes the error structure of the whole (μ_v and σ_v^2).

Two important rules for the propagation of errors can now be stated. Firstly, the expected value of a sum of random values is equal to the sum of their expected values. Thus, if V = X + Y,

(A2) $$E[V] = E[X + Y] = E[X] + E[Y]$$

or

(A2′) $$\mu_v = \mu_x + \mu_y .$$

In our example, the presumption of zero mean errors in the estimates of wages and profits implies that the estimate of national income is also distributed with zero mean error.

The second and more important rule concerns the dispersion of errors. Once again, let V = X + Y. Then (A1) may be rewritten:

(A3) $$\sigma_v^2 = E[(X-\mu_x)+(Y-\mu_y)]^2$$

(A3′) $$= E[(X-\mu_x)^2+(Y-\mu_y)^2+2(X-\mu_x)(Y-\mu_y)]$$

Using (1) above, and noting that $(X-\mu_x)(Y-\mu_y) = \mathrm{cov}(X,Y) = \sigma_{xy}$, (3′) may be replaced by the more elegant variant:

(A4) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy},$$

where σ_{xy} indicates the covariance of the two terms, and measures the extent to which the two variables share common patterns over repeated samplings of their joint (bivariate) distribution. If higher values of X are systematically related to higher values of Y, then their covariance is positive; if the tendency of the two variables is to move in opposite directions, their covariance is negative.[48]

Thus, the variance of a linear combination of variables is equal to the sum of the variances of the component parts, plus their covariance. The corresponding standard error of V, σ_v, is derived by taking the square root of the variance:

(A5) $$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}}$$

48 A third important rule, which we shall not prove here, is that, if the error structures of *X* and *Y* are assumed to be normally distributed, then the error of the whole, *V*, will also be normally distributed, with means and variances as defined in (2) and (4) above.

It is possible to add one further variant to the statistical treatment of errors, which will be useful in later discussion, namely that, since σ_{xy} is by definition equal to $r_{xy}\sigma_x\sigma_y$, where r_{xy} is the correlation coefficient of x and y, (A4) and (A5) may be rewritten as:

(A6) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 + 2r_{xy}\sigma_x\sigma_y$$

and

(A7) $$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + 2r_{xy}\sigma_x\sigma_y} \ .$$

Relationships between the errors

We must now take into account the relationship between the errors in the component series. If we assume that all errors are independent, so that X and Y are not systematically related to each other, the covariance, σ_{xy}, (or the correlation coefficient, r_{xy}) is set equal to zero, and (A4) and (A6) both collapse to:

(A8) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 .$$

This represents the general rule of the additivity of variances, which states that the variance of a linear combination of independent variables is equal to the sum of the variances of the individual variables. This is the most important basic result for our analysis of the propagation of errors. The corresponding standard error of the total is:

(A9) $$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2} \ .$$

If, however, the errors are not independent (because, for example, the estimates are drawn from a common source), it is necessary to estimate the extent of correlation (r_{xy}) and proceed accordingly. Three stylized examples can be given to illustrate the effects of dropping the assumption of independence:

(i) If the errors in the estimates of wages and profits are perfectly (positively) correlated (being drawn from the same imperfect sample of company accounts, perhaps), then $r_{xy} = 1$, and (A6) becomes:

(A10) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y$$

which collapses to:

(A11) $$\sigma_v^2 = (\sigma_x + \sigma_y)^2$$

such that:

(A12) $$\sigma_v = \sigma_x + \sigma_y$$

In this case, the total error is additive not in the variances, but in the standard errors, of the parts.

(ii) If we assume perfect negative correlation between the errors (as would arise from subtracting total wages from overall company net output to estimate profits, for example), then $r_{xy} = -1$, and (A6) becomes:

(A13) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y$$

which collapses to:

(A14) $$\sigma_v^2 = (\sigma_x - \sigma_y)^2$$

such that:

(A15) $$\sigma_v = \sigma_x - \sigma_y .$$

Interdependence of construction implies that we should not double-count the errors in the two components.

(iii) Finally, if the errors are partially correlated with each other, such that r_{xy} =, say, 0.5, (A6) becomes:

(A16) $$\sigma_v^2 = \sigma_x^2 + \sigma_y^2 + \sigma_x\sigma_y ,$$

such that:

(A17) $$\sigma_v = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_x\sigma_y} .$$

This is a common procedure followed when it is thought that the two error terms are not wholly independent of each other, but are not sufficiently interrelated to require the assumption of perfect correlation.[49] There is, of course, nothing sacrosanct about the valuation of 0.5 for r_{xy}; researchers may apply whatever value they think appropriate between – 1 and + 1.

The basic rules can be applied to the error structure of any variable constructed as the sum of a series of independent estimates. Thus, in Bowley's pioneering work on the earnings of 'intermediate' income recipients (those who neither earned wages nor paid income taxes) in pre-1914 Britain, the error of the entire estimate was generated by summing the variances of the total earnings of each 'intermediate' occupation (i.e. neither wage-earners nor tax-payers).

[49] See, for example, Chapman, *Wages and Salaries*, p. 236.

In all such cases, the total value of an estimate, N, is the sum of the value of the parts, N_i, (for example, the income in each intermediate occupation). The formula given in (A8) for two parts can now be generalized to cover any number of parts. Since the error of the total is equal to the sum of the errors of the parts, we have:

(A18) $$V = \Sigma V_i$$

and

(A19) $$\sigma_v^2 = \Sigma \sigma_{vi}^2$$

where σ_{vi} is defined as the standard error of V_i. The variance of the sum is equal to the sum of the variance of the parts, a result that follows logically from the rule of additive variances.

We can now introduce the concept of subjective margins of error. In the 95% probability version, the proportionate standard error, $\varepsilon_{vi} = \frac{\sigma_{vi}}{N_i}$, is defined as half the assigned proportionate margin of error (ε_{vi}). An alternative expression for the variance of the parts on the right hand side of (A19) is thus:

(A20) $$\sigma_{vi}^2 = \varepsilon_{vi}^2 . N_i^2$$

This establishes the required relationship between the margin of error as a subjective concept, designed to organize thoughts about the reliability of an estimate, and the standard error as a formal statistical concept.

The calculus of errors

Many estimates are, however, more complicated in form than a simple aggregated series. Thus, most value terms in economics may be decomposed into price and quantity components. The wage-bill in any industry is the product of average wages and total employment in that sector; consumer expenditures are the product of quantities purchased and average price. The procedure for dealing with this common case is simplified by invoking elementary calculus, as follows.

Let F be a function of the form:

(A21) $$F = f(X, Y) .$$

The error structure of F, denoted by ΔF, will depend upon the errors of the component parts of F. The error structure of the function may be derived by total differentiation of $f(.)$. Thus, (A22) $$\Delta F = \frac{\partial F}{\partial X}\Delta X + \frac{\partial F}{\partial Y}\Delta Y$$

where $\partial F / \partial X$, &c., are the partial derivatives of F with respect to X, &c. In general calculus, $\Delta Y\ (\partial Y)$, &c. tend to zero at the limit. Our approach, however, applies to discrete changes, derived from the underlying probability distribution of errors. We may simplify the notation by replacing $\partial F / \partial X$ by F_X and ΔX by σ_X and similarly for the other terms. Doing this, and squaring (A22) produces the following relationship between the variance of the components and of the whole:

(A23) $$\sigma_F^2 = (F_x\sigma_x)^2 + (F_y\sigma_y)^2 + 2(F_x F_y \sigma_x \sigma_y r_{xy})$$

where r_{xy} is the correlation of σ_X and σ_Y.

Despite the slight change in symbols, it is clear that this functional form is symmetric to (A6) above, which we derived using the rules of probability, rather than calculus. Once again, if we assume independence of the errors (such that the correlation between X and Y is zero), (A23) collapses to:

(A24) $$\sigma_F^2 = (F_x\sigma_x)^2 + (F_y\sigma_y)^2$$

which is equivalent to (A8) above.[50]

This general approach may be applied to the propagation of errors from any functional form. The most common forms are set out in Table A1. To assist in the understanding of the approach, let us explicate two of these variants, the product and the quotient.

a. *The product*

Let F be defined as the product of two terms, w and X, where X is a vector of values and w is a vector of weights. Thus, the average manufacturing wage is the weighted average of the wage in each manufacturing industry (X_i), where the weights are the share of total employment in each sector (w_i).[51] This may be expressed as:

(A25) $$F = \sum_i w_i X_i .$$

[50] Note that if $F = X + Y$, $\partial F/\partial X = F_x = 1$ and $\partial F/\partial Y = F_y = 1$.

[51] This example is presented with relative weights (i.e. $\Sigma w_i = 1$); it makes no difference to the analysis if the weights are absolute.

The error in F depends upon the error in the weights (w_i) and values (X_i) respectively. Applying standard differentiation rules:

(A26) $$\delta F = \Sigma\left[\left(\frac{\partial F}{\partial X}\delta X_i\right) + \left(\frac{\partial F}{\partial w}\delta w_i\right)\right]$$

(A27) $$\delta F = \Sigma\left[(w_i \delta X_i) + (X_i \delta w_i)\right]$$

which holds for small changes in the variables.

We can express (A27) in probabilistic terms by replacing δF, &c. by δ_F, &c., and squaring, such that

(A28) $$\sigma_F^2 = \Sigma[w_i^2 \sigma_{xi}^2 + X_i^2 \sigma_{wi}^2 + 2(w_i X_i \sigma_{xi} \sigma_{wi} r_{xw})]$$

where σ_F is the standard error of F, and only correlations between $x_{i \neq j}$ and $w_{i \neq j}$ are considered. It follows that, under the governing assumption of the independence of errors, the variance of the weighted average is:

(A29) $$\sigma_F^2 = \Sigma w_i^2 \sigma_{xi}^2 + X_i^2 \sigma_{wi}^2$$

This in turn may be rewritten for ease of interpretation as:

(A30) $$\sigma_F^2 = \Sigma w_i^2 X_i^2 \left(\frac{\sigma_{xi}^2}{X_i^2} + \frac{\sigma_{wi}^2}{w_i^2}\right)$$

such that the variance of the weighted average is equal to the weighted sum of the squared coefficients of variation of the two variables, where the weights are the squared value F^2 in each sector, *i*.

Invoking (A20), and noting that $w_i^2 X_i^2 = F_i^2$, it can easily be shown that:

(A31) $$\frac{\sigma_v^2}{F^2} = \Sigma(\varepsilon_{xi}^2 + \varepsilon_{wi}^2)\frac{F_i^2}{F^2}$$

which states that the variance of the total is equal to the weighted sum of the proportionate variances of its component parts. This form provides a more intuitively appealing approach to the problem of the propagation of errors.

b. *The quotient*

The rules applying to a quotient, noted in Table 1, may be developed for a number of variants beyond the simple case of x / y. Among these are the ratio of two weighted averages (such as the average cost of living at two dates, where each observation is the weighted average of individual prices), and the decompounded growth rate of a variable or index between any two points in time. We concentrate here on the case of a ratio of two weighted averages, viz.

(A32) $$F = \frac{\Sigma w_i X_i}{\Sigma y_i Z_i}$$

where y_i, w_i are the proportionate weights of sector i in periods 1 and 2, respectively; and Z_i, X_i are the average values of sector i in periods 1 and 2.

Invoking (A24) and, for the sake of simplicity, assuming independent errors for now, we may write:

(A33) $$\sigma_F^2 = \Sigma(F_w\sigma_{wi})^2 + (F_X\sigma_{Xi})^2 + (F_y\sigma_{yi})^2 + (F_Z\sigma_{Zi})^2$$

where F_w symbolises $\partial F / \partial w$, &c. The partial derivatives take the following form:

$$F_w = \sum(X_i / \Sigma y_i Z_i); \qquad F_x = \sum(w_i / \Sigma y_i Z_i)$$

$$F_y = \sum[(w_i X_i) Z_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i + \sigma_{yi} Z_i)] \qquad F_z = \sum[(w_i X_i) y_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i + \sigma_{zi} y_i)].$$[52]

Once again, for ease of interpretation, the expression may be rewritten, by dividing both sides of (A33) by F and simplifying:

$$F_w / F = (\Sigma X_i / \Sigma y_i Z_i)/(\Sigma w_i X_i / \Sigma y_i Z_i) = (\Sigma X_i / \Sigma y_i Z_i) \times (\Sigma y_i Z_i / \Sigma w_i X_i) = \Sigma X_i / \Sigma w_i X_i$$
$$F_x / F = (\Sigma w_i / \Sigma y_i Z_i)/(\Sigma w_i X_i / \Sigma y_i Z_i) = (\Sigma w_i / \Sigma y_i Z_i) \times (\Sigma y_i Z_i / \Sigma w_i X_i) = \Sigma w_i / \Sigma w_i X_i$$
$$F_y / F = (\Sigma(w_i X_i) Z_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i \pm \sigma_{yi} Z_i) / (\Sigma w_i X_i / \Sigma y_i Z_i)$$
$$= (\Sigma(w_i X_i) Z_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i \pm \sigma_{yi} Z_i) \times (\Sigma y_i Z_i / \Sigma w_i X_i) = (\Sigma Z_i / (\Sigma y_i Z_i \pm \sigma_{yi} Z_i)).$$
$$F_z / F = (\Sigma(w_i X_i) y_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i \pm \sigma_{Zi} y_i) / (\Sigma w_i X_i / \Sigma y_i Z_i)$$
$$= (\Sigma(w_i X_i) y_i / (\Sigma y_i Z_i)(\Sigma y_i Z_i \pm \sigma_{Zi} y_i) \times (\Sigma y_i Z_i / \Sigma w_i X_i) = (\Sigma y_i / (\Sigma y_i Z_i \pm \sigma_{Zi} y_i)).$$

Thus, the proportionate error in the ratio is:

(A34)
$$\left(\frac{\sigma_F}{F}\right)^2 = \Sigma\left(\frac{X_i}{\Sigma w_i X_i}\sigma_{wi}\right)^2 + \Sigma\left(\frac{w_i}{\Sigma w_i X_i}\sigma_{Xi}\right)^2 + \Sigma\left(\frac{Z_i}{\Sigma y_i Z_i \pm \sigma_{yi} Z_i}\sigma_{yi}\right)^2 + \Sigma\left(\frac{y_i}{\Sigma y_i Z_i \pm \sigma_{Zi} y_i}\sigma_{Zi}\right)^2$$

The proportionate error in the weighted average is the sum of the variances of each term, suitably weighted.

This formula is complete if the weights are independent (as would be the case if each w_i were, say, the number of workers in each wage category and the total number of workers was not fixed). However, if the weights in the index are proportional (such that the weights sum to one), the formula needs to be revised to incorporate the necessary accounting constraints. This is necessary because of the interdependence of the weights. Thus, if the weight, w, in sector i is too

---

[52] Note the peculiarity of this derivation; the standard result for F_y suppresses the term, ΔZ_i, in the denominator, since δZ_i tends towards its limit of zero. However, in the finite case of errors, $\sigma_{Z\iota} > 0$, and should be retained in the complete expression.

high, there must be compensating changes in the weights attached to the remaining sectors, given $\Sigma w_i = 1$. The most plausible procedure for distributing the cross-effects involves proportional changes in the weights of the remaining sectors. In this case, F_{wi} should be rewritten:

$$\text{(A35)} \qquad \frac{\partial F}{\partial w_i} = \frac{\sum X_i - \sum (w_j / \Sigma w_j) X_j}{\sum y_i Z_i}$$

where $i \neq j$. Similarly for F_{yi}:

$$\text{(A36)} \qquad \frac{\partial F}{\partial y_i} = \frac{\sum w_i X_i [Z_i - (\sum \frac{w_i}{\Sigma w_j} Z_j)]}{\sum y_i Z_i - [\sum y_i Z_i - (Z_i - \sum \frac{w_i}{\Sigma w_j} X_j) \sigma_{yi}]}$$

The other partial derivatives in equation (A33) remain the same.

Finally, we relax the assumption of independent errors. In general, we follow the procedure laid down in (A23) above, in which the covariance between terms are captured by $F_x F_y \sigma_x \sigma_y r_{xy}$, for each combination, where r_{xy} can take positive or negative values between 0 and 1. In the case of a weighted average, there are ten pairwise combinations ($r_{ww}, r_{xx}, r_{yy}, r_{zz}, r_{wx}, r_{wy}, r_{wz}, r_{xy}, r_{xz}, r_{yz}$), assuming symmetry. In many cases, the pairwise correlation will be zero (in the case, for example, of independently derived estimates of weights for the base- and end-years); in these cases, the additional term vanishes. Note that the complexity of the formulation for a weighted average derives from the fact that the denominator and numerator are themselves combinations of estimates with non-negative covariances, in which weights and values may also be cross-correlated.

There is no need to rehearse the entire set of possible pairwise combinations. The formula is the same for each case, combining the appropriate partial derivatives, standard errors and the assumed correlation coefficient. Thus, in the case of the interdependence between measurement of X_i in the denominator, the formula will be:

$$r_{xx} \quad : 2\left(\frac{w_i \sigma_{xi} . w_j \sigma_{xj}}{\Sigma y_i Z_i . \Sigma y_i Zi}\right) r_{xixj} \quad \text{for pairs of } i, j \text{ for } i, \ldots, n, \text{ where } i \neq j$$

$$r_{xx} \quad = \left(\frac{w_i \sigma_{xi} . w_j \sigma_{xj}}{\Sigma y_i Z_i . \Sigma y_i Zi}\right) r_{xixj} + \left(\frac{w_i \sigma_{xi} . w_j \sigma_{xj}}{\Sigma y_i Z_i . \Sigma y_i Zi}\right) r_{xjxi} \quad \text{for each member of a pair}$$

(e.g. for cotton yarn and cotton cloth, where the first term refers to cotton cloth, and the second to cotton cloth). A similar construction ($F_Z F_Z \sigma_Z \sigma_Z r_{ZZ}$) will be appropriate for pairs of i and j in the numerator. The other formulae follow this same template.

Table 1
The structure of errors in the Hoffmann index of industrial production.
(percentage margins of error)

Sector	1780–1801				1801–1831			
	Indicator		Value added		Indicator		Value added	
	1780	1801	1783	1812	1801	1831	1812	1850
Coal mining	10	10	50	50	10	10	50	25
Copper ore	10	10	50	50	10	10	50	25
Iron & steel	50	25	75	75	25	25	75	25
Iron goods, &c.	50	50	50	50	50	50	50	25
Copper	10	10	50	50	10	10	50	25
Tin	25	25	50	50	25	25	50	25
Copper goods	50	25	75	75	25	25	75	25
Ships	10	10	50	50	10	10	50	25
Furniture	50	50	75	75	50	50	75	25
Other timber goods	50	50	75	75	50	50	75	25
Cotton yarn	10	10	50	50	10	10	50	25
Cotton cloth	--	25	--	50	25	25	50	25
Wool yarn	25	25	50	50	25	25	50	25
Wool cloth	25	25	50	50	25	25	50	25

/... continued overleaf

Continued from previous page

Silk yarn	10	10	50	50	10	10	50	25
Silk cloth	25	25	50	50	25	25	50	25
Hemp	--	10	--	50	10	10	50	25
Linen	50	25	50	50	25	25	50	25
Flour	50	50	75	75	50	50	75	25
Bread	50	50	75	75	50	50	75	25
Sugar	10	10	50	50	10	10	50	25
Beer	10	10	50	50	10	10	50	25
Malt	10	10	50	50	10	10	50	25
Spirits	50	10	75	75	10	10	75	25
Tobacco	10	10	50	50	10	10	50	25
Paper	10	10	50	50	10	10	50	25
Leather	50	10	75	75	10	10	75	25
Leather goods	50	25	75	75	25	25	75	25
Soap, candles	50	10	75	75	10	10	75	25
Misc. industries	50	50	75	75	50	50	75	25
Clothing, &c.	50	50	75	75	50	50	75	50

Notes: The indicator for Cotton Yarn in 1780 includes both Cotton Yarn and Cotton Cloth. There is no indicator for Hemp in 1780.

Table 2: Deriving the error margins for the Hoffmann index of industrial production, 1780–1831

		Fully independent		Not independent	
		(1)	(2)	(3)	(4)
I: 1780–1801					
Terms:	w_i	0.0337	0.0005	0.0337	0.0005
	X_i	0.0135	0.0135	0.0135	0.0135
	Z_i	0.0168	0.0168	0.0168	0.0168
	y_i	0.0432	0.0002	0.0432	0.0002
	r_{xx}			0.0015	0.0015
	r_{zz}			0.0040	0.0040
	r_{ww}			–0.0058	–0.0058
	r_{yy}			–0.0116	–0.0116
Variance of ratio		0.1072	0.0310	0.0952	0.0190
SE of ratio		0.3275	0.1761	0.3085	0.1377
Ratio 1801/1780		2.20	2.20	2.20	2.20
Error ±		29.8 %	16.0 %	28.1 %	12.5 %
Range		1.54 to 2.81	1.84 to 2.55	1.58 to 2.85	1.92 to 2.47
Growth rate (% p.a.)		3.82	3.82	3.82	3.82
Range		2.08 to 5.12	2.96 to 4.56	2.20 to 5.05	3.16 to 4.40
II: 1801–1831					
Terms:	w_i	0.0182	0.0007	0.0182	0.0007
	X_i	0.0239	0.0239	0.0239	0.0239
	Z_i	0.0158	0.0158	0.0158	0.0158
	y_i	0.0396	0.0004	0.0396	0.0004
	r_{xx}			0.0014	0.0014
	r_{zz}			0.0016	0.0016
	r_{ww}			–0.0018	–0.0018
	r_{yy}			–0.0066	–0.0066
Variance of ratio		0.0.974	0.0408	0.0920	0.0354
SE of ratio		0.3122	0.2020	0.3033	0.1881
Ratio 1801/1780		2.29	2.29	2.29	2.29
Error ±		27.2 %	17.6 %	26.5 %	16.4 %
Range		1.67 to 2.92	1.89 to 2.70	1.69 to 2.90	1.92 to 2.67
Growth rate (% p.a.)		2.80	2.80	2.80	2.80
Range		1.72 to 3.63	2.14 to 3.36	1.76 to 3.61	2.19 to 3.33

Note: Columns (1) and (3) make no adjustment for compensating weights; columns (2) and (4) incorporate the adjustment.

Table A1: Basic formulae for the propagation of errors

	$\sigma_F^2 =$
F	$(F_x\sigma_x)^2 + (F_y\sigma_y)^2 + 2(F_xF_y\sigma_x\sigma_y r_{xy})$
$x + y$	$\sigma_x^2 + \sigma_y^2 + 2(\sigma_x\sigma_y r_{xy})$
$x - y$	$\sigma_x^2 + \sigma_y^2 - 2(\sigma_x\sigma_y r_{xy})$
xy	$(y\sigma_x)^2 + (x\sigma_y)^2 + 2(yx\sigma_x\sigma_y r_{xy})$
x/y	$\left(\frac{\sigma_x}{y}\right)^2 + \left(\frac{x\sigma_y}{y^2}\right)^2 - 2\left(\frac{x}{y^3}\sigma_x\sigma_y r_{xy}\right)$

BIBLIOGRAPHY

Berg, Maxine and Pat Hudson. "Rehabilitating the Industrial Revolution." *Economic History Review* 45 (1992): 24–50.

Bowley, Arthur L. "Changes in Average Wages in the United Kingdom between 1880 and 1891." *Journal of the Royal Statistical Society* 58 (1895): 223–85.

— — — . "The Statistics of Wages in the United Kingdom during the last Hundred Years. Part IV, Agricultural Wages, Earnings and General Averages." *Journal of the Royal Statistical Society* 62 (1899): 555–70.

— — — . "The Measurement of the Accuracy of an Average." *Journal of the Royal Statistical Society* 75 (1911–12): 77–88.

— — — . *Elements of Statistics*. 6th ed. London: P. S. King, 1937.

— — — . "National Income in America and the United Kingdom." *Economica* 9 (1942): 27–36.

Bowley, Arthur L. and Josiah Stamp. *The National Income 1924*. Oxford: Clarendon Press, 1927.

British Association. "Report of the Committee (A. L. Bowley, Secretary) on the Amount and Distribution of Income (other than Wages) below the Income Tax Exemption Limit in the United Kingdom." *Journal of the Royal Statistical Society* 74 (1910–11): 37–66.

Britton, Andrew and David Savage. "The Three Measures of GDP." *National Institute Economic Review* 107, no. 1 (1984): 54–9.

Central Statistical Office. *National Accounts Statistics, Sources and Methods*. London: HMSO, 1968.

Chapman, Agatha, L. *Wages and Salaries in the United Kingdom 1920–1938*. Cambridge: Cambridge University Press, 1953.

Clark, Colin. *The National Income, 1924–1931.* London: Macmillan, 1932.

— — — . *National Income and Outlay*. London: Macmillan, 1937.

Cole, Roseanne. *Errors in Provisional Estimates of Gross National Product*. New York: NBER, 1970.

Crafts, N.F.R. *British Economic Growth during the Industrial Revolution*. Oxford: Clarendon Press, 1985.

Crafts, N.F.R., Leybourne, S.J. and Mills, T.C. "Trends and Cycles in British Industrial Production," *Journal of the Royal Statistical Society*, Ser. A. 152 (1989): 43–60.

Crafts, N.F.R. and Harley, C.K. "Output Growth and the British Industrial Revolution: a Restatement of the Crafts-Harley View." *Economic History Review* 45, no. 4 (1992): 703–30.

Cuenca Esteban, J. "British Textile Prices, 1770–1831: are British Growth Rates Worth Revising Once Again?" *Economic History Review* 47, no. 1 (1994): 66–105.

— — — . "The Rising Share of British Industrial Exports in Industrial Output, 1700–1851." *Journal of Economic* History 57, no. 4 (1997): 879–906.

Deane, Phyllis and W. A. Cole. *British Economic Growth, 1688–1959*. Cambridge: Cambridge University Press, 1962.

Deane, Phyllis. *The Measurement of Colonial National Incomes*. NIESR Occasional Paper XII. Cambridge: Cambridge University Press, 1948.

Deming, W. Edwards. *Statistical Adjustment of Data*, New York: John Wiley, 1943.

Desai, R. C. "Consumer Expenditure in India, 1931–2 to 1940–1." *Journal of the Royal Statistical Society* 111, pt III (1948): 261–98.

Fabricant, Solomon. *The Output of Manufacturing Industries, 1899–1937*. NBER. Princeton: Princeton University Press, 1940.

Federico, Giovanni and Antonio Tena. "On the Accuracy of Foreign Trade Statistics (1909–1935): Morgernstern Revisited." *Explorations in Economic History* 28, no. 3 (1991): 259–73.

Feinstein, Charles H. *Domestic Capital Formation in the United Kingdom, 1920–1938*. Cambridge: Cambridge University Press, 1965.

— — — . *National Income, Expenditure and Output of the United Kingdom, 1855–1965*. Cambridge: Cambridge University Press, 1972.

Feinstein, Charles H. and Katherine Watson. "Private International Capital Flows in Europe in the Inter-war Period." In *Banking, Currency and Finance in Europe between the Wars,* edited by Charles Feinstein, 94–130. Oxford: Oxford University Press, 1995.

Gallman, Robert E. "Gross National Product in the United States, 1834–1909." In *Output, Employment and Productivity in the United States after 1800,* edited by Dorothy S. Brady, 3–76, NBER Studies in Income and Wealth, vol. 30. Princeton: Princeton University Press, 1966.

— — — . "The United States Capital Stock in the Nineteenth Century." In *Long-Term Factors in American Economic Growth,* edited by Stanley L. Engerman and Robert E. Gallman, 165–213. NBER Studies in Income and Wealth, vol. 51. Chicago: Chicago University Press, 1986.

Gilbert, Milton, ed. *Studies in Income and Wealth*, vol III. Cambridge: Bowes and Bowes, 1953.

Goldsmith, Raymond W. *A Study of Saving in the United States*. 3 vols. Princeton: Princeton University Press, 1955.

Greasley, David and Les Oxley. "Balanced versus Compromise Estimates of UK GDP, 1870–1913." *Explorations in Economic History* 32, no. 2 (1995): 262–72.

Harley, C. K. "British Industrialization before 1841: Evidence of Slower Growth During the Industrial Revolution." *Journal of Economic History* 42, no. 2 (1982): 267–90.

Harley, C. K. and N. F. R. Crafts. "Cotton Textiles and Industrial Output Growth during the Industrial Revolution." *Economic History Review* 48, no. 1 (1995): 133–44.

Hoffmann, W. G. *British Industry, 1700–1950.* Oxford: Blackwell, 1955.

Hoppit, Julian. "Counting the Industrial Revolution." *Economic History Review* 43, no. 2 (1990): 173–93.

Jackson, Robert V. "Rates of Industrial Growth during the Industrial Revolution." *Economic History Review* 45, no. 1 (1992): 1–23.

Kendrick, John W. *Productivity Trends in the United States*. NBER Princeton: Princeton University Press, 1961.

Kuznets, Simon. *National Income and its Composition, 1919–1938*. 2 vols. New York: NBER, 1941.

— — — . *Capital in the American Economy*. NBER Princeton: Princeton University Press, 1961.

Mitchell, Wesley, C., W. I. King, et al. *Income in the United States, its Amount and Distribution*. Vol. 1. New York: Harcourt Brace,1921.

Morgernstern, Oskar. *On the Accuracy of Economic Observations*. 2nd revised ed. Princeton: Princeton University Press, 1963.

Prest, Alan R. assisted by Arthur Adams. *Consumers' Expenditure in the United Kingdom, 1900–1919.* Cambridge: Cambridge University Press, 1954.

Rowe, J. W. F. "The Physical Volume of Production." London and Cambridge Economic Service, Special Memorandum, no. 8. London, 1924.

Roy, Andrew D. "An Exercise in Errors." *Journal of the Royal Statistical Society* 115, pt. IV (1952): 507–20.

Sefton, James A. and Martin Weale. *The Reconciliation of National Income and Expenditure: Balanced Estimates of National Income for the United Kingdom, 1920–1990*. Cambridge: Cambridge University Press, 1995.

Solomou, Solomos and Martin Weale. "Balanced Estimates of UK GDP, 1870–1913." *Explorations in Economic History* 28, no. 1 (1991): 54–63.

Stone, Richard. "National Income in the U.K. and the U.S.A." *Review of Economic Studies* 10 no. 1 (1942–3): 1–27.

— — — . “Two Studies on Income and Expenditure in the United States.” *Economic Journal* 53, no. 1 (1943), 60–75.

Stone, Richard, D. G. Chanpernowne and James E. Meade. “The Precision of National Income Estimates.” *Review of Economic Studies* 9, no. 2 (1942): 111–25.

Temin, Peter. “Two Views of the British Industrial Revolution.” *Journal of Economic* History 57, no. 1 (1997): 63–82.

United States, Department of Commerce. *National Income Supplement, 1954*.

Weale, Martin. “The Reconciliation of Values, Volumes and Prices in the National Accounts.” *Journal of the Royal Statistical Society* Series A 151 pt 1 (1988): 211–21.

Weiss, T. J. “U.S. Labor Force Estimates and Economic Growth, 1800–1860.” In *American Economic Growth and Standards of Living before the Civil War*, edited by Robert E. Gallman and J. J. Wallis, 19–78. Chicago: Chicago University Press, 1992.

Williamson, Jeffrey G. *Did British Capitalism Breed Inequality?* London: George Allen and Unwin, 1985.

[Continued from inside front cover]

26 David Stead, *An Arduous and Unprofitable Undertaking: The Enclosure of Stanton Harcourt, Oxfordshire* (November 1998)

27 Oliver Grant, *The Diffusion of the Herringbone Parlour: A Case Study in the History of Agricultural Technology* (December 1998)

28 Antonia Taddei, *London Clubs in the Late Nineteenth Century* (April 1999)

29 Liam Brunt, *Estimating English Wheat Production in the Industrial Revolution* (June 1999)

30 Matthew Braham, *Volunteers for Development: A Test of the Post-Materialism Hypothesis in Britain, c. 1965–1987* (June 1999)

31 Paul A. David and Gavin Wright, *General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution* (September 1999)

32 Liam Brunt, *An Arbitrage Model of Crop Rotation* (September 1999)

33 Paul A. David and Gavin Wright, *Early Twentieth Century Productivity Growth Dynamics: An Inquiry into the Economic History of 'Our Ignorance'* (October 1999)

34 Avner Offer, *Economic Welfare Measurements and Human Well-Being* (January 2000). Rev. version, March 2000.

35 Liam Brunt, *'Where there's Muck, There's Brass'. The Market for Manure in the Industrial Revolution* (February 2000).

36 Alasdair Crockett, *Variations in Churchgoing Rates in England in 1851: Supply-Side Deficiency or Demand-Led Decline?* (August 2000).

37 Martin West, *State Intervention in English Education, 1833–1891: A Public Goods and Agency Approach* (October 2000).

38 George Speight, *Who Bought the Inter-War Semi? The Socio-Economic Characteristics of New-House Buyers in the 1930s* (December 2000)

39 Peter Temin, *A Market Economy in the Early Roman Empire* (March 2001)

40 Michael Biggs, *Positive Feedback in Collective Mobilization: The American Strike Wave of 1886* (April 2001)

41 Charles H. Feinstein and Mark Thomas, *A Plea for Errors* (July 2001)

University of Oxford
Discussion Papers in
Economic and Social History

are edited by:

Papers may be downloaded from
http://www.nuff.ox.ac.uk/Economics/History/