

### Statistical Sources in France before World War I

*Claude Diebolt\**

**Abstract:** Today, quantitative data are doubtless quite rightly occupying an increasingly large position in economic history. We are no longer in the period of vague descriptions and collections of isolated facts that did not lead to any valid explanation. The influence of economists involved historians in the handling of figures, essential measures for those who wish to understand structures and detect movements. It is nonetheless true that there are serious differences between the approach of economists and that of historians. Economists apply reasoning to practically only the present time or to a relatively short period. Their models and patterns are difficult to apply to periods when the structures were markedly different. They also use regular statistical series covering a considerable number of facts, and above all series that may not be perfect (perfection is illusory here) but provide serious guarantees. Historians are less privileged. They dissect economic systems that are very different to our own and whose structures have not yet been closely studied. They possess only sparse statistics whose reliability seems extremely doubtful. Economists are not usually faced with the problems of source and critique that are the daily lot of historians.

This article is aimed less at providing complete results than stimulating certain research on detail. Our statistical knowledge and our knowledge of statistical data are still too fragmentary and imperfect for it to be possible to envisage

---

\* Address all communications to: Claude Diebolt, BETA/CNRS, Université Louis Pasteur de Strasbourg, Faculté des Sciences Economiques et de Gestion, 61 Avenue de la Forêt Noire, 67085 Strasbourg Cedex, France, e-mail: [cdiebolt@cournot.u-strasbg.fr](mailto:cdiebolt@cournot.u-strasbg.fr).

an absolutely definitive overall study. The reader will find here only the components of a general problematic. The main reasons for this attitude are explained below.

At a time when group work is becoming commonplace, it seems important to establish solid initial foundations. A whole set of laboratory work must be set in motion. Only this will provide us with the material required for developing valid historical economics. And this technique can doubtless be best used in the field of statistics. There is no laboratory without instruments—instruments that are changed gradually with research, that are ceaselessly perfected.

The problem of sources is itself extremely complex. Before addressing the material aspect, its nature should perhaps be defined more clearly. The terms ‘statistics’ encompasses fairly varied notions and causes confusion. Statistics, as we use it now, is in the great majority of cases figures recorded administratively, accounts kept regularly in accordance with known rules. The material is created, developed and perfected as much by administrative necessity as by certain scientific requirements. The more we go back in time, the fewer the series defined in this way, and they are doubtless less accurate. The chronological limits of statistics thus vary.

To mention a few examples, the boundaries are 1770 for population movement, 1819 for foreign trade, 1825 for crimes and offences and 1833 for the mineral industry. I shall only go into detail on the subject of education.

In education, our specialist field here, a statistical commission for primary education was instituted for the first time by the French Ministry of Public Instruction by an order of 15 March 1876. Its aim was to publish detailed statistics concerning primary education every five years. It was chaired from the beginning by Émile Levasseur and consisted of senior Ministry of Public Instruction officials and representatives of the *Conseil supérieur de statistique*.

Eight volumes were published before World War I. The first, covering the 1876-1877 academic year, was an explicit continuation of the 5-year population census to facilitate comparison of the number of children to be educated and that of the children enrolled or attending schools. The second volume was by far the largest (more than 700 pages), entitled *Statistique comparée de l'Enseignement Primaire de 1829 à 1877* (Comparative Statistics of Primary Education from 1829 to 1877). It reports practically all the statistics available for this period. The other volumes are for the 1881-1882, 1886-1887, 1891-1892, 1896-1897, 1901-1902 and 1906-1907 academic years. The sixth volume (1896-1897) contains a summary of the progress in primary education under the Third Republic. The last volume was published at the end of 1909. The documents intended for the ninth volume (1911-1912) were assembled on the eve of World War I but publication was not possible. The statistical commission no longer functioned after this date. After the 1914-1918 war, *Statistique Générale de la France* collected the most important data on primary teaching

on an annual basis, using original Ministry of Education records (number of schools, classes, schoolmasters and pupils by department, etc.). It is nevertheless unnecessary to underline the inadequacy of such documentation in comparison with the previous 5-year publications.

In addition, part of the statistics collected but not published is housed in national or departmental archives. Consultation of these makes up for the lack of scholastic statistics from 1914 to 1958 (much data was collected during this period but a comparatively small proportion was published, probably for budgetary reasons). In a general manner, statistics concerning the education system in France in the nineteenth and twentieth centuries can be classified under two main headings. On the one hand, it consists of statistical data produced by one-off investigations performed at the initiative of a minister, a special commission or a group of members of parliament to examine certain aspects of the functioning of educational facilities. For example, the figures include statistics on primary, upper primary, secondary and higher education. They also comprise information collected during the everyday management of educational institutions or of education as a whole and whose main (and sometimes sole) purpose was administrative. This concerns mainly data published in the statistical yearbooks of the SGF (*Statistique Générale de la France*), the main basis for most of the overall studies of the trends in school attendance in France.

Historians were certainly not satisfied with these limits and therefore naturally envisaged the reconstitution of certain data series. Given the state of the documentation available, it would indeed have been impossible to extend the technique to a large number of sectors. However, what is possible today, what is feasible for a short perspective is much more difficult for long periods in the past. Statistical reconstitution can be envisaged in a limited number of cases with satisfactory approximation. It is desirable when the aim is more to be able to define a trend than provide statistical details. However, it is often impossible. Here again, we can examine a few examples. The history of prices, apart from fearsome difficulties in currencies and measures, has abundant sources in certain continuous accounts. Applying sampling methods to registry data is perfectly justified with regard to statistical method and can provide interesting information for the appraisal of a population. This is unfortunately not the case for other sectors such as wage rates or the discount rate for example, or for industrial production and many other areas.

We do not intend to address the difficult, fascinating problem of constituted statistics. Research techniques are all different and would deserve a special study—very strongly wished for and, in fact, begun. The scope of the present work is limited to existing statistics, true administrative statistics. We felt that it was essential to first draw up a sincere balance of what already exists, precisely to avoid as much as possible the need for always delicate reconstitution work. Other distinctions are just as important. The problem of the scope of statistical information seems essential first of all. In a national framework, statisticians

today work on comparatively unified sets of data. Historians are faced with an extremely disparate universe that makes overall series suspect. This aspect of statistics is very well illustrated historically. Before the second half of the nineteenth century, attention was paid less to continuous data series than to broader public surveys capable of revealing structures that do not appear with figures alone. This is more than a simple statistical comment. It seemed essential to detail the specific features of fractionation to which we are less and less accustomed today.

Our notion of statistics is therefore extremely broad. We consider that the major surveys that opened up broad perspectives from the time of Colbert onwards are necessarily within our subject. Their existence clearly shows that at least as much importance was awarded to structures as to trends. The very texts of the surveys and the questions asked are ample proof that they were not only the result of fear of regular questioning. They were not only an easy solution to the problem.

All this leads us directly to another major question, that of the degree of statistical information. The distinction between overall figure and partial figure is important. It is all the more important as it operates in two sectors of prime importance—time and space. Historians and economists are unanimous here in desiring as much fractionation of figures as possible.

Although at first sight they appear easier to solve, the problems of time nevertheless raise difficulties that should not be misunderstood. It is not necessary here to mention the calculation of means and indexes; these concern statistical techniques alone and assume, of course, the availability of data running over a short period of time. It is precisely this short period that is involved here. Monthly statistics are common today (price indexes, activity indexes, etc.). They used to be rarer, for reasons that are easy to understand.

Under the 'Ancien Régime', only market price lists were produced for short periods (a fortnight). In the nineteenth century, quite a long time passed before regular monthly population movement statistics were instituted. Indeed, it is fairly simple to make a logical determination of the frequency of statistics. It is very difficult to re-establish statistics for shorter periods than those drawn up by the administration, except when there is continuous documentation, as for population or prices. The problem should be underlined nonetheless to prevent the destruction of documentation.

The main preoccupation in the question of space is certainly the same, and even more acute. The basic figures of all statistics are necessarily local and sometimes even smaller than the scale of the parish or the commune. Louis-Philippe's industrial statistics were compiled at the scale of businesses. Agricultural statistics are generally compiled at commune level. Recapitulations for larger administrative zones are frequent but do not make the basic figures any less interesting. Population movement statistics under the 'Ancien Régime' were drawn up by parish and generally only known for general features (at

least they were published in this form or sometimes at an intermediate administrative level). Parish statistics, which were in fact the only type to provide effective monitoring, remain of primordial interest. I shall return to this question shortly.

Research in statistics is thus a very broad field. It does not stop naturally at economic data alone and does not handle overall figures only. The only limit found was that of the conservation of documents. A rapid examination of the resources of archives and libraries in France would doubtless lead to considerable pessimism. Much has been destroyed and present conservation conditions do not appear to be totally satisfactory. The first observation to be made is that all the statistics generally resulted from government initiatives. All or almost all should therefore have been centralised in Paris. Hardly anything of the efforts of the 'Ancien Régime' governments remains in the archives in Paris. The archives of the '*Contrôle général*', the key body in all these surveys, have disappeared. What was handed over to the ministries that inherited its functions represents very little from our point of view. However, we know that everything was centralised. Not much is left—one or two large surveys and a few crumbs of regular information.

The hesitations of nineteenth century governments with regard to statistics no doubt considerably affected the conservation of documents. The '*Service de la Statistique Générale de la France*' (SGF) that was the heir in 1833 of an imperial department that had disappeared under the Restoration, did store documents. The material that it deposited in the national archives is extremely fragmentary. Most of the documents concern the work that led to publications and is not therefore of much real scientific interest. The situation is thus paradoxical as much of what has been kept consists of the tables published by the SGF and almost everything else is missing.

Local archives thus remain our only substantial source of material. This is all the more important as it is only there that figures for the smallest administrative units can be found. And it is also only there that statistical historians can find the additional documents to support solid critique—reminder letters, explanatory letters, statistical commission reports, successive progress reports, etc. It is at this level that the work of a research laboratory devoted to statistical history should truly start. The situation is very varied in this field. Over and over again the destruction caused by successive wars and that seriously affected certain geographical administrative departments, many disappointments lay in wait for historians. Documents of this type are placed in the M section of departmental archives, a key section for contemporary history and which is too large and too disorganised. These documents suffered disastrous sorting before and after filing that resulted in the loss of a fair proportion of the survey of wages during the July Monarchy, work which is obviously of great interest.

Searching is difficult in itself for want of inventories, and often for want of classification. We do not even have as many as ten printed inventories for this

section and elsewhere the lists are often inadequate. Moreover, classification has been too neglected and logical searches are impossible. The list of blunders and ignorance is endless: surveys are mixed up (the 1840 listing of fiscal questions is often found with population censuses, post-1852 agricultural surveys are set in with 10-year surveys), dispersed (the 1840-1844 industrial survey is often divided into several parts elsewhere and care was not taken to separate the questionnaires with different dates). Efforts to achieve the logical grouping of the valid components of the documents have been rare. The only documents that have received a degree of respect seem to be the population censuses and market prices.

The situation is just as bad for documents dating from the 'Ancien Régime' (we have mentioned that they are hardly to be found at all in Paris). It is also noted that the documents are very badly entered in inventories and it is very possible that our lists are incomplete. We dearly hope that an effort has been made in classification, at least for the modern period. In spite of these deficiencies and gaps, rich documentation does exist that is likely to provide good material for research. But historians and economists go separate ways here again. The latter are used to homogeneous, continuous data series in which the risks of error, although still present, are limited. The situation is completely different for older documents. Indeed, past statisticians had no illusions with regard to their work, often criticising its inadequacy or even its basic inaccuracy. Historians should not lose their critical sense when faced with precise, dry notation that is easily dominant. This is certainly not the occasion for writing a treatise of statistical criticism. We still possess too little material for this. Partial research has nevertheless shown that such criticism is possible.

Regular statistics with a long past are certainly much more likely to be accurate than an episodic survey. We have clear proof of this. From the beginnings of statistics on population movements in 1773, the government has taken care to monitor the slightest details and indicate the most perfect methods. The administration's desire for statistical series that are as accurate as possible can be felt clearly. Long, real supervision of the work necessarily gives satisfactory results. In contrast, the risk of error is much greater in episodic surveys and the work carried out at the time of Louis-Philippe and the Second Empire farming surveys seem to be fairly remote from the real situation. It was also desired that the surveys should be as broad-ranging as possible and the persons handling the replies were often submerged by the questions. Insofar as it performed statistical surveys the 'Ancien Régime' operated discreetly in limited, divided fields that required only comparatively simple tables. The Consulate and the Empire were partially responsible for these questionnaires easily extending to many hundred pages or numbers. These episodic surveys, and this can be seen in relation to the 1862 agricultural survey, gained firmer foundations insofar as regular statistics were developing, leading to a kind of statistical education for everybody.

So these are several initial criteria for a regular start to statistics and whose routine brings accuracy by a kind of psychological process that is easy to re-constitute whereas episodic statistics encourage all kinds of irregularities. On the one hand more or less than observed before with rigour, and on the other an overall view that can only be overall through lack of accuracy. There are other, just as simple distinctions. A statistical object is much easier to enter when it has already been recorded. Therein lies all the difference between industrial or trade statistics and agricultural statistics, between the statistics for foreign trade monitored by the customs authorities and the movement of domestic trade. The only limit to accuracy in this field is that opposing private accounting and public accounting. From the mid-eighteenth century, and even more so throughout the nineteenth century, fear of tax adjustments was a serious danger for statistical accuracy. At the local scale, it should be relatively easy to perform simple criticism; requests for explanations—those of the person who has trouble understanding a question and those of the administration that has trouble understanding a result—are there to provide information and guide us. At the central scale, criticism can only be that of abstract logic made up of concomitance and inexplicable deviations. It therefore has less value and also less effectiveness.

This elementary critical work was already performed at the time. The question of statistical control is a difficult one and was a problem for a long time. The issue of the nature of this control has been discussed since the nineteenth century. Should it be purely administrative or collective and almost democratic? The statistical commissions outlined after the 1836-1839 agricultural survey and taken up again in 1848 and made systematic in 1852 attempted to set up continuous control of the people concerned. It was certainly a failure, as shown by all the documents.

Several forms of *a posteriori* control are possible. It exists at the local scales, insofar as the administrative personnel has close knowledge of the local situation. Marc Bloch was full of admiration for subdelegates' personnel who were very aware of the problems in their areas and knew how to give ratings to reports and responses that would have been difficult to obtain elsewhere. It is clear that prefects in the nineteenth century, who remained in the same post for a long time, also had close knowledge of their departments. These administrators could in any case always call on local worthies capable of appraising the results of such surveys. This is seen later on the subject of the 1839 agricultural survey in the Côtes-du-Nord department. The results of the 10-yearly agricultural surveys performed under the Second Empire were submitted to Chambers of Agriculture, some of which seem to have made interesting rectifications.

Logical control and scientific control can also be effective and it is the latter that functions best today. Both reveal excessive disparities indicating poor methods in the market reports of the first half of the nineteenth century. A number of surveys were performed to study the methods used in various fields

and to achieve stricter control in the market reports under Louis-Philippe and in population movements in 1855 under Napoleon III. These controls became stricter as the century progressed, that is to say to the extent that progress was made in the science of statistics.

It can therefore be seen that the historian's task is facilitated by pre-existing critique, much of which is still present in the files. Naturally, none of this is in the printed publications except in certain introductions that outlined comparisons with previous surveys.

There is no question here of discussing statistical methods as these are covered in a respectable number of specialised works. At most, mention can be made of a few questions that are preliminaries to any serious study. Before any calculation, historians of statistics must ask themselves about the homogeneity of the sets of data. For example, geographical homogeneity assumes breaks at the moment of administrative reforms. To mention only one case, the parishes of the 'Ancien Régime' do not correspond to the modern communes in some regions. In prices, changes in quality can be mentioned. French statistics thus give the wholesale price of coal with two quality breaks, in 1832 and 1840. However, it would be possible to assemble homogeneous data sets for a clearly specified period of time. The problem of quality comes down to the statistical expression of the evolution of techniques, and this is not easy.

The internal homogeneity of figures can be a more difficult problem. The explanation is that statistical habit leads us, or should lead us, to increasing accuracy. It can therefore be considered that the grain harvest figures provided annually by the French national statistical body are more accurate as time goes by. Contemporaries had understood this well, considering that appraising harvests in the first half of the nineteenth century was not very serious as cadastral mapping operations had not been completed. It is also possible that this statistical uncertainty may have affected the sensitivity of staples markets, as estimations of production and consumption should be treated with great caution. And it is doubtless here that attentive study of sources enables us to identify zones of certainty besides immense areas of uncertain truth. Although agricultural statistics seem to be of very poor quality, even according to the persons collecting and using them, mining industry statistics and judicial statistics appear on the other hand if not models of the kind at least satisfactorily completed work. We even reach the paradox, that is not a reproach for the personnel of SGF, that major administrative statistics are of much better quality than the work of SGF. With the exception of several areas such as hospital statistics or those of charity, the latter had the task of performing all the research that could not be carried out by administrations that did not have the necessary powers or the internal regulations essential for conducting this work.

Another important question of interpretation arises beyond the question of the gradual, logical perfecting of statistical knowledge. It is that regarding the changes of the habits of the population and alternative solutions. Precise cases

can be mentioned. We possess rail traffic statistics since the beginning of the railways. It is difficult and even impossible to make a distinction in these figures between what is a variation in the economic situation and what is the growth of a new type of transport replacing others concerning which we have no figures.

Likewise, the discount figures of a public bank, the Banque de France or departmental Banks, reflect both the growth or recession of the economy and the development of the institution itself, the confidence gradually placed in it by the public and the development of branches supplying discountable material. The figure may contain not one but several different movements that cannot be separated. In other words, some statistics reveal not only the characteristics of the situation but also the evolution of certain structures.

Insofar as the figure revealing an overall rather than a partial phenomenon that cannot incorporate the idea of the replacement of another phenomenon as the figure is an index of movement only, it sometimes gives us only a pale image of fluctuation. The data series in our possession enable us to glimpse real facts rather than the intentions that strongly reflect oscillations and sometimes make it easier to understand certain phenomena.

Two examples explain this serious problem.

Halbwachs considered that the number of suicides is a fairly accurate index of the amount of suffering, anxiety, imbalance and sadness existing or produced in a group. The suicide rate has been known in France through two different sets of statistics, at least for a certain length of time now. These are legal statistics and statistics of the causes of death drawn up for population movements. The first source mingles suicides and attempted suicides while the second counts only effective suicides. But it is obvious that the number of attempted suicides is distinctly under-estimated in the first category, to the extent to which the difference between the two figures gives data for attempted suicides that have no relation with reality. The records were only started for Paris in 1956 and with certain reserves.

Steel factories are unhealthy places and subject to administrative authorisation. This is one of the bases of statistics for the industry. Here, we have not just one figure but a whole series of figures, some of which are more interesting than others. We can first discard the purely administrative angles. The authorisation may indeed arrive several years after the application, several years after the construction of the works and even when a slump halts operation. The administrative series of authorisations is therefore often aberrant and cannot therefore be used. A series of controls by engineers who count the production machines and monitor production itself is a real data series. The curve drawn from it is valid but although the dips mark the low points in slumps the peaks mark less well the desires for expansion in scale and also in time. Here again, there are both attempts and successes, as for suicide. We only have information about the successes, but the administration can also provide us with

information about attempts; the statistics for applications for authorisation gives us a measure of the hopes and also the dates—approximate of course given the studies performed and the slow administrative procedures—of entrepreneurs' decisions. Historians may thus possess a means of analysis with regard the special area of entrepreneurs' decisions. The metals industry is certainly a special sector in this respect. There are others, however. The forming of the railway companies in 1845-1847, which received abundant press coverage, makes it possible to observe much larger movements of capital than hitherto believed. There may therefore be material for interesting research and fresh explanations. Statistics still has unexplored zones.

A whole book would be necessary to give all the details of critiques and methods required for the good use of old statistics and to which economists today are not accustomed. The aim here was to provide a few glimpses. A full study requires long laboratory work and attentive, extended practice.

### Selected References

- CHAULEUR, A. (1980) *Bibliothèques et archives: comment se documenter?*, Economica, Paris, 334 pages.
- DESROSIERES, A. (1993) *La politique des grands nombres. Histoire de la raison statistique*, La Découverte, Paris, 441 pages.
- DIEBOLT, C.; JAOL, M.; SAN MARTINO, G. (2003) "Primary Education by Administrative Department in France in the 19th Century. A Cliometric Analysis in Terms of Convergence", *American Economic Association—Allied Social Science Associations—Economic History Association*, Session on Education and Intergenerational Transfers Across Time and Space, Washington, D.C., 3-5 January 2003, 36 pages.
- [Http://www.cliometrie.org](http://www.cliometrie.org).