Chapter 3

Backward Calculation Techniques

3.1 Introduction

Data currently produced by National Statistical Offices (NSIs) are often subject to a revision process. The revision process can be viewed according to its double dimension: routine revisions and occasional revisions. For both of them the main purpose is to achieve better quality of the published data. While the former are regularly made to incorporate the new available information in order to improve the quality of the statistics, the latter occur at irregular intervals depending on major accounting events.

Occasional revisions are produced by NSIs at longer and infrequent intervals. The nature of such revisions may be statistical, when results from changes in surveys or in estimation procedures, or conceptual, when results from changes in concepts, definitions or classifications. The effect of an occasional revision increases according to the interval that occurs between two successive revisions. Clearly, if the intervals become longer, the effects of each revision become larger creating more difficulties both for the accountants and the users. This implies also more work in revising, and consequently, more difficulties in managing the revision process.

NSIs often face with occasional revisions. For example, when they change the reference year for constant price figures, they entail a revision process that involves the entire national accounting framework. From a conceptual point of view, occasional revisions are different because generated by different causes. Causes of this type might be

- Changes due to new surveys
- Changes due to modifications in definitions or interpretations of the European System of Accounts (ESA) and the System of National Accounts (SNA)
- Introduction of new calculation methods

- Important economic events that have a big impact on the national accounting system. An example of such an economic event is the introduction of EURO.

These occasional revisions are out of the usual scheme of revision and ask for a deep analysis of the impact they have on the national accounting system and of the strategy that accounts should follow to implement them. The main effect of these revisions is to affect all national accounts. Time series associated to the national accounts aggregates have to be revised according to the new changes. As a result of this process, national accountants and econometricians want to have their disposal national accounting series that are homogeneous and at the same time cover the longest possible period. The reconstruction of the national accounts time series is associated to a revision process usually referred as backward calculation or retropolation.

Eurostat´s Unit B2 is now revising the backward calculation methodologies adopted by some Member States. The reason for that is because Eurostat wants to make suggestions to other Member States that have not developed methodology on the backward calculation techniques to develop one. This is very important especially at this time that there is an increasing demand for homogeneous time-series inside the European Union. Furthermore, Eurostat foresees the development of the same methodology internally. It is apparent that the role of Eurostat in this estimation domain aims at the over space harmonization of the backward calculation methods.

The present chapter starts with the description of specific areas in which backward calculation is necessary. Next we give an analytical description of the most well known backward calculation methods (annual backward calculation method and benchmark years and interpolation method). The Netherlands and the France methodological approaches are described thoroughly focusing more on theoretical aspects of the Kalman-Filter method used by France. In the final part some concluding remarks about the retropolation methods are given.

3.2 The introduction of EURO, The 1993 System of National Accounts Regulation and The 1995 European System of Accounts Regulation. Three cases Where Backward Calculation is Required

3.2.1 Backward Calculation Due to The Introduction of EURO

The Problem

Until 31st December 1998, Member States of the Euro-zone, as well as the remaining EU Member states, sent to Eurostat their statistical data expressed in national currency. To obtain the figures for the European aggregates and to publish them, these data were converted in ECU according to the exchange rate quoted on financial markets. Time series associated to national accounts variables were then stored in the commission databases both in national currency and in ECU.

Starting from the 1st January 1999, the national currencies of the Member States of the Euro-zone and EURO are both official currencies. National currency will be valid until 2002 and Member states will continue to publish their national accounts time series in national currency until the end of 2001. Possibly, Member States will switch to EURO in their national accounts before the scheduled deadline or they will publish both figures. In any case, during the period January 1999-December 2001 national currency and EURO will live together.

In order to have homogeneous time series, during the transitioning period, Member States will convert their historical time series, expressed in national currency, in euro series. This operation corresponds to a backward calculation of the involved time series. The main difficulties in this problem arise from the fact that ECU and euro do not represent the same thing. Officially, starting from 1st January 1999 the following relation stands.

1 euro = 1 ECU

However, for the period preceding the euro birth, the problem of the conversion of national currencies in ECU or euro arises.

Solution

According to a recent note signed by Commissary De Silguy, the following decisions have been taken on this subject:

- National currency historical time series will be converted in euro according to the exchange rate fixed on 30th December 1998 (the official exchange rates).

- Historical time series previously expressed in ECU will not change. Starting from 1st January 1999, figures will be expressed in euro. Euro time series will be the statistical continuation of the ECU time series. A sort of break will characterize the label of these time series: until 31st December 1998 they will be expressed in ECU and afterwards in euro. The users of official statistics will be informed by suitable footnotes.

- Two types of historical time series will describe Euro-zone national accounts variables: the Ecu-euro time series, as described just previously and the new fixed euro time series obtained by converting the old national currency series according to the euro exchange rate fixed on 30th December 1998. The main aim of these ″fixed″ time series is to avoid contrasts between historical time series produced by Member States and Eurostat.

From a statistical point of view the ECU-euro time series and the ″fixed″euro time series are different because of the fixed nature of the exchange rate between euro and national currency in contrast to the floating rate of ECU and national currency. The two different natures of the series will imply different use and interpretation of the statistics associated to them.

3.2.2 Backward Calculation Due to The Introduction of ESA[13] 1995 Regulation and The SNA[14] 1993 Regulation

The SNA Problem

The 1993 System of National Accounts states that the System of National Accounts consists of a coherent, consistent and integrated set of macroeconomic accounts, balance sheets and tables based on a set of internationally agreed concepts, definitions, classifications and accounting rules. One of the strong points of the SNA is that it creates a basis not only for detailed snapshots but especially for comparisons over time. In practice the accounts are compiled for a succession of time periods, thus providing a continuing flow of information that is indispensable for the monitoring, analysis and evaluation of the performance of the economy over time. Time-series of national accounting data give a detailed account of the economic development of a country over time. Most countries in the world compile national accounts according to the same international guidelines, which makes possible the comparison of economic developments between those countries.

Although the time dimension is mentioned in the SNA, nothing is said about the method to compile time-series. In fact, the SNA gives guidelines for forward calculation of national accounting data and not for backward calculation. In the 1993 SNA (and in the 1995 ESA) the words revision, time-series, backward calculation or retropolation does not appear. Furthermore, despite the fact that revisions of international guidelines are one of the important motives to revise national accounts and to obtain time-series which are consistent with the revised data for the revision year, these revisions and the time series are not discussed. The fact that the session of the twenty-third General Conference of the International Association for Research in Income and Wealth (New Brunswick, Canada,1994) was devoted to policies for revisions of national accounts is an indication that revisions and time-series are considered to be a problem with no clear answers.

---

[13] European system of accounts.
[14] System of national accounts.

The fact that the international guidelines do not discuss revisions and backward calculations of national accounting data is probably one of the reasons that different revision policies occur in different countries. This concerns differences with regard to the frequency of the revision, the choice of the revision date, the level of the detail at which the benchmark year is revised, the method for compiling time-series and the length and detail of the time series. Differences in methods between countries obscure comparisons between countries. As a result in order to achieve an international comparability, the Member States of the European Community made some arrangements for the backward calculation of national accounting data.

The ESA Problem

Starting form April 1999, Member States will publish national accounts figures according to the European System of Accounts. The introduction of European System of Accounts regulation entails several changes in the compilations of national accounts. These changes demand for a backward calculation of time series to ensure coherence and homogeneity of the series describing national accounts aggregates.

The introduction of European System of Accounts regulation is another typical example of occasional revision due to changes in definitions and in interpretation. It is a major occasional event that asks for a deeply analysis of the impact on the national account system and on the historical national accounts time series. Because of the importance of the changes that the European System of Accounts regulation introduces in the national accounts systems, many Member States already analyzed the effects of the new accounting rules and in some cases already started the reconstruction of the time series.

In the following sections, several techniques of backward calculation of national accounts time series proposed by some Member States are analyzed and discussed in order to suggest to the remaining Member States concrete strategies to carry out backward calculation.

3.3 An Overview of Backward Calculation Techniques

One of the main aims of national accounts is to create a base for over time comparison of economic variables. In fact European System of Accounts (ESA) and System of National Accounts (SNA) regulation does not give any guidelines concerning backward calculation or retropolation. For that reason, different solutions have been adopted by Member States in order to treat this problem.

Generally speaking, in the methods for backward calculation of national accounting data two archetypes can be distinguished

- **Annual backward calculation**
- **Benchmark years and interpolation**

In both methods several variants are possible and also a combination of both methods is thinkable. For example in the Netherlands case a number of variants of the first class of methods were used in the past. Until now, the second class was not used except from the revision of the national accounting data in the interwar period. The former class of methods is very well known in National Statistical Institutes and methods belonging to them are currently used to revise time series. The latter has not been intensively applied till now to revise national accounts series.

3.3.1 Annual Backward Calculation

Annual backward calculation is based on the principle that the retrapolated figures are calculated year by year back in time. Several methods can be used to obtain such results. The differences among them depend more or less on accuracy, and consequently time used in carrying out the revision process and more or less on the intensive use of statistical techniques. The most well known methods belonging to the backward calculation class are the following:

Full Revision Method

The full revision method is a very complete one. Figures to be revised, covering all the years in the backward calculation period, are estimated by applying the same

principles that underlie the revision. This means that in the case of the application of European System of Accounts regulation, past years are estimated according to the new rules established by that regulation. This procedure due to its detailed level of analysis, asks for the existence of a very good system of basic statistics suitable to be re-used according to the new classifications and revisions. Clearly, this method is very time consuming and requires much staff.

Revision by Superposition of Corrections

Time series figures concerning the years of the backward calculation period are determined by superposing corrections on the figures before revision. Starting point is the consistent data set of national account which was compiled in the past. Corrections resulting from the revision process are added to this basic set. The revision process involves all the past years.

Two cases can be distinguished when applying this method: The former corresponds to a superposition of a set of corrections already calibrated on the complete accounting context; the latter implies the revision of the concerned item, the extension of the revision of the concerned items to all periods and the consolidation of all accounts.

Growth Rates Method

Starting from the balanced set of national accounts figures for the revision year, time-series figures for the past are determined by applying backwards the growth rates associated to the time series before revision. Obviously, if revised growth rates for a certain variable are available, they are used. The revision process works at the level of detail chosen. Afterwards, the figures are balanced again in the framework of a consistent national accounts system.

Simple Proportional Method

The simple proportional method is a simplified version of the annual backward calculation method. The revision year is expressed both under the new and the old

accounting system rules. Then in order to reconstruct the past revised values of the series, a simple proportional rule is applied to the old time series values.

The simple proportional method offers an easy technique to carry out backward calculation, especially in a first attempt to determine the new path of the involved time series. Clearly, it is an approximate solution that does not analyze in a very deep way the revision effects on time-series but on the contrary is a low resource and time consumption approach to the backward calculation.

3.3.2 Benchmark Years and Interpolation

The second group of basic methods for backward calculation of national accounts is based on a two step procedure. In the first step detailed estimates for one or more benchmark years are calculated. In the second step, figures for the remaining years are determined by interpolation. The benchmark years and interpolation method can be applied in different ways i.e. the full benchmark year method and the layer correction method

The Benchmark Years

Before starting the present description we have to give some basic definitions. As we have already said retropolation of national accounting data is necessary after a revision of the national accounts has taken place. Such a revision is carried out for the so-called revision year. Consequently the revision year is that one for which the new definitions and accounting rules are used for first time. The new figures for that year are determined at a very detailed level using the new accounting rules. Revision years and benchmark years are strongly connected. Actually the revision year is an outstanding example of a benchmark year and is the starting point for the backward calculation of the data. It is obvious that the benchmark years are crucial points in the time series and they should include as much information as possible. That's why benchmark years are usually years in which population, occupation or industrial censuses are conducted. Furthermore the economic situation is of great importance for the choice of the benchmark years. The corrections which are carried out for the revision year have to be determined for the other benchmark years as well. After a

number of revisions have been carried out in due time for all benchmark years, strata of corrections matrices are available (one for each revision).

Interpolation

After the revision corrections for the benchmark years have been determined, the corrections for the intermediate years are calculated by interpolation. However, the method of interpolation differs between Member States. In the next section we are going to analyze the estimation approaches for the intermediate years as well as the history of backward calculation of two pioneering countries (in the field of backward calculation); the Netherlands approach and the France approach.

Benchmark years interpolation method seems to be the solution to the problem of backward calculation according to the methodologies proposed by Member States in this field. This method has a number of advantages.

- The method is transparent and relatively fast,
- The revision corrections are determined explicitly,
- Decisions are taken in the past in the balancing of the data are upheld
- In the case of new revision only the revision corrections for the benchmark year have to be determined.

As far as it concerns the interpolation method is flexible, not to much time consuming and it can use direct information about certain variables for one or more years.

Full Benchmark Year

Figures for the benchmark years are estimated in a detailed way, using new definitions, classifications and sources. The benchmark years should be very well known and should dispose a complete set of basic information The integration and balancing of benchmark years is made according to the new accounting rules. This means that the integration decisions taken in the past, are not used. The figures for the years between the benchmark years are interpolated. The level of detail is chosen according to the level of detail of the revision in the benchmark years and is extended to the interpolated years. In the interpolation process the historical tracks of previous revisions are taken into account.

Layer Correction

Figures for the benchmark years are determined starting from the original, balanced data set. Corrections, resulting from revisions, are balanced and then superposed to the basic set of original data. In this way, layers of correction matrices become available (one for each revision). Corrections for the intermediate years are determined by means of interpolation of the correction matrices. Afterwards, figures for the years between the benchmark years are determined by integrating the original data and the corrections.

According to this method, corrections are determined for all years in the period to be revised. However, not all years are treated in the same way. Especially in estimating the revision corrections the difference between benchmark years and other years is evident. The figures for the benchmark years are estimated with the help of detailed information. The figures for the other years are estimated more roughly.

3.4 The Netherlands Case

In the Netherlands there is a tradition of almost 60 years in compiling time-series of national accounts data. In 1948 the first revision of national accounting data (referring to the years 1921-1939) was published.

In the Netherlands the various methods for backward calculation have been used extensively. The annual backward calculation has been used for compiling time-series 1977-1985, consistent with the 1987 revision. The revision by superposition of corrections method was used for compiling the time-series 1969-1977, following the 1977 revision. The growth rates method was used for compiling the time-series 1969-1976 following the 1987 revision. The simple proportional method was originally planned for the time-series 1948-68 according to the 1987 revision. The full benchmark year method was originally planned for compiling time-series for the years 1969-1986 following the 1987 revision. However, in the end the annual backward calculation method and especially the growth rates method were used.

Recently, Statistics Netherlands decided to change the method of backward calculation. The new method is a variant of the layer correction method belonging to the benchmark years and interpolation category.

In the Dutch case the corrections for the intermediate years are calculated with the help of an interpolation procedure that has been developed for this purpose (Kazemier, 1997). The interpolation is carried out within the framework of the input-output tables.

To start with, the corrections for the benchmark years are expressed as a percentage of the values before the correction. This means, if we take a column of the input-output table as an example, that for each cell of the column a correction percentage is determined. These percentages are interpolated between the benchmark years by assuming a linear pattern. Next, the percentages are applied to the values before the corrections. In this way the before revision structures are kept and the developments of the variables before revision are taken into account.

The system for the interpolation is very flexible. For instance, it is possible to use direct information about a certain variable for one or more years. The corrections are determined in such a way that for all non-benchmark years integrated sets of corrections are obtained. These sets have the same structure and level of detail as those for the benchmark years.

In the Dutch situation the present benchmark years are 1987,1977,1969,1958 and 1948. The first two years are revision years. For instance in the reporting year 1969 the 1968 SNA was implemented. In that year a reclassification of enterprises was carried out and the value added tax was introduced. The year 1958 is chosen because in the past extra attention was paid in that year in the compilation of the national accounts. The year 1948 is a benchmark year because it is the first year for which the international guidelines of the 1953 SNA and the 1952 OECD were implemented. In addition, the 1948 is the first `normal` year after the Second World War.

3.5 The French Case

The French approach to the estimation of the intermediate years is based on a linear model that links the variable in the new accounting system and the variables of the accounting system before revision. The estimates of this linear model are obtained by applying the Kalman –filter algorithm. Consequently, before focusing on the French's exact methodology we will describe the Kalman-Filter algorithm.

3.5.1 The Kalman Filter

In 1960, R.E. Kalman published his famous paper describing a recursive solution to the discrete-linear filtering problem. The idea behind Kalman´s work is to express a dynamic system in a particular form called state-space representation.

A Kalman–Filter is simply an optimal recursive data processing algorithm. There are many ways of defining optimal, depending upon the criteria chosen to evaluate the performance. One aspect of this optimality is that the Kalman-Filter incorporates all information that can be provided to it. It processes all available measurements, regardless of their precision, to estimate the current value of the variables of interest, with use of (1) knowledge of the system and the measurement device dynamics (2) the statistical description of the system of noises, measurement noises and uncertainty in the dynamics of the models and (3) any available information about initial conditions of the variables of interest. The word recursive in the previous description means that the Kalman-Filter does not require all previous data to be kept in storage and reprocessed every time a new measurement is taken. A Kalman-Filter combines all available measurement data, plus prior knowledge about the system and measuring devices, to produce an estimate of the desired variables in such a manner that the error is minimized statistically. In other words, if we were to run a number of candidate filters many times for the same application, then the average results of the Kalman-Filter would be better than the average results of any other. Conceptually, what any type of filter tries to do is to obtain an "optimal" estimate of desired quantities from data provided by a noisy environment.

The State-Space Representation of a Dynamic System - Maintained Assumptions

Let y_t denote an $(n \times 1)$ vector of variables at date t. A rich class of dynamic models for y_t can be described in terms of a possibly unobserved $(r \times 1)$ vector ξ_t known as the state vector. The state space representation of the dynamics of y is given by the following system of equations:

$$\xi_{t+1} = F\xi_t + v_{t+1} \quad (3.5.1.1)$$

$$y_t = A'x_t + H'\xi_t + w_t \quad (3.5.1.2)$$

where F, A' and H' are matrices of parameters of dimension $(r \times r), (n \times k), (n \times r)$, respectively and x_t is a $(k \times 1)$ of exogenous or predetermined variables. Equation (3.5.1.1) is known as the state equation while equation (3.5.1.2) is known as the observation equation. The $(r \times 1)$ vector v_t and the $(n \times 1)$ vector w_t are white noise vectors:

$$E(v_t v_\tau') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \quad (3.5.1.3)$$

$$E(w_t w_\tau') = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \quad (3.5.1.4)$$

where Q and R are $(r \times r)$ and $(n \times n)$ matrices respectively. The disturbances v_t and w_t are assumed to be uncorrelated at all lags:

$$E(v_t w_\tau') = 0 \text{ for all } t \text{ and } \tau \quad (3.5.1.5).$$

The statement that x_t is predetermined or exogenous means that x_t provides no information about ξ_{t+s} or w_{t+s} for s=0,1,2,… beyond that contained in $y_{t-1}, y_{t-2}, ..., y_1$. Thus, for example, x_t could include lagged values of y or variables that are uncorrelated with ξ_t and w_t for all τ .

The system (3.5.1.1) through (3.5.1.5) is typically used to describe a finite series of observations $\langle y_1, y_2, ..., y_T \rangle$ for which assumptions about the initial value of the state vector ξ_t are needed. As a result we assume that ξ_t is uncorrelated with any realizations w_t or v_t:

$$E(v_t \xi_1') = 0 \text{ for } t = 1,2,...,T \quad (3.5.1.6)$$
$$E(w_t \xi_1') = 0 \text{ for } t = 1,2,...,T \quad (3.5.1.7)$$

Thus, (3.5.1.6) and (3.5.1.3) imply that v_t is uncorrelated with lagged values of ξ_t:

$$E(v_t \xi_\tau') = 0 \text{ for } t = t\text{-}1, t\text{-}2,...1. \quad (3.5.1.8).$$

Similarly,

$$E(w_t \xi_\tau') = 0 \text{ for } \tau = 1,2,...,T \quad (3.5.1.9)$$
$$E(w_t y_\tau') = E[w_t(A'x_\tau + H'\xi_\tau + w_\tau)'] = 0 \text{ for } \tau = t-1, t-2,...1. \quad (3.5.1.10)$$
$$E(v_t y_\tau') = 0 \text{ for } \tau = t-1, t-2,...1. \quad (3.5.1.11)$$

The system (3.5.1.1) through (3.5.1.7) is quite flexible, though it is straightforward to generalize the results further to systems in which v_t is correlated with w_t.

Derivation of The Kalman Filter

Consider the general state-space:

$$\xi_{t+1} = F\xi_t + v_{t+1}$$

$$y_t = A^{'} x_t + H^{'}\xi_t + w_t$$

$$E(v_t v_\tau{'}) = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

$$E(w_t w_\tau{'}) = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases}$$

The analyst is presumed to have observed $\langle y_1, y_2, ..., y_T, x_1, x_2, ... x_T \rangle.$ One of the ultimate objectives may be to estimate the values of any unknown parameters in the system on the basis of these observations. Actually there are many uses of Kalman-Filter. It is motivated here as an algorithm for calculating linear least squares forecasts of the state vector on the basis of the data observed through date t,

$$\hat{\xi}_{t+1/t} = \hat{E}(\xi_{t+1} / \Psi_t) \text{ where}$$
$$\Psi_t = (y_t{'}, y_{t-1}{'}, ..., y_1{'}, x_t{'}, x_{t-1}{'}, ..., x_1{'}){'} \text{ (3.5.1.12)}$$

and $\hat{E}(\xi_{t+1} / \Psi_t)$ denotes the linear projection of ξ_{t+1} on Ψ_t and a constant. The

Kalman-Filter calculates these forecasts recursively, generating $\hat{\xi}_{1/0}, \hat{\xi}_{2/1}, ..., \hat{\xi}_{T/T-1}$ in succession. The mean squared error (MSE) associated with each of these forecasts, is represented by the following $(r \times r)$ matrix:

$$P_{t+1/t} = E\left[(\xi_{t+1} - \hat{\xi}_{t+1/t})(\xi_{t+1} - \hat{\xi}_{t+1/t}){'} \right] \text{(3.5.1.13)}$$

Starting The Recursion

The recursion begins with $\hat{\xi}_{1/0}$ which denotes a forecast of $\hat{\xi}_1$ based on no observations y or x. This is just the unconditional mean of ξ_1, $\hat{\xi}_{1/0} = E(\xi_1)$ with associated mean square error given by the following formula:

$$P_{1/0} = E\left\langle \left[\xi_1 - E(\xi_1) \right]\left[\xi_1 - E(\xi_1) \right]' \right\rangle .(3.5.1.14)$$

A general rule for the initialization of the algorithm can be the following. Provided that the eigenvalues of F matrix are all inside the unit circle, then the process for ξ_t in (3.5.1.1) is covariance stationary. The unconditional mean of ξ can be found by taking expectations of both sides of (3.5.1.1), producing

$E(\xi_{t+1}) = FE(\xi_t)$, or since ξ_t is covariance stationary, $(I_r - F)E(\xi_t) = 0$.

Since unity is not an eigenvalue of F, the matrix $(I_r - F)$ is nonsingular, and this equation has unique solution $E(\xi_t) = 0$. The unconditional variance of ξ can be similarly found by postmultiplying (3.5.1.1) by its transpose and taking expectations:

$E(\xi_{t+1}\xi_{t+1}') = E\left[(F\xi_t + v_{t+1})(\xi_t'F' + v_{t+1}')\right] = FE(\xi_t\xi_t')F' + E(v_{t+1}v_{t+1}')$. Cross product terms have been disappeared in the light of (3.5.1.8). Letting Σ denote the variance covariance matrix of ξ, this equation implies that $\Sigma = F\Sigma F' + Q$ whose solution by theory is given by $\text{vec}(\Sigma) = \left[(I_{r^2} - (F \otimes F)\right]^{-1}\text{veq}(Q)$. Thus in general in the initialization step of the algorithm we can focus in the following rules.

If the eigenvalues of F are inside the unit circle then the Kalman-Filter iterations can be started with $\hat{\xi}_{1/0} = 0$ and $P_{1/0}$ the $(r \times r)$ matrix whose elements expressed as a column vector are given by $\text{vec}(P_{1/0}) = \left[(I_{r^2} - (F \otimes F)\right]^{-1}\text{veq}(Q)$.

If instead some eigenvalues of F are on or outside the unit circle, or if the initial state ξ_1 is not regarded as an arbitrary draw from the process implied by (3.5.1.1), then

$\hat{\xi}_{1/0}$ can be replaced by analyst's best guess as to the initial value of ξ_1, where $P_{1/0}$ is a positive definite matrix that summarizes the confidence about this guess. Larger values for the diagonal elements of $P_{1/0}$ register greater uncertainty about the true value of ξ_t.

Forecasting y_t

Given starting values $\hat{\xi}_{1/0}$ and $P_{1/0}$ the next step is to calculate analogous magnitudes for the following date, $\hat{\xi}_{2/1}$ and $P_{2/1}$. The calculations for $t = 2,3,...,T$ all have the same basic form, so we will describe them in general terms for step t ; given $\hat{\xi}_{t/t-1}$ and $P_{t/t-1}$ the goal is to calculate $\hat{\xi}_{t+1/t}$ and $P_{t+1/t}$. First, since we have assumed that x_t contains no information about ξ_t beyond that contained in Ψ_{t-1},

$$\hat{E}(\xi_t / x_t, \Psi_{t-1}) = \hat{E}(\xi_t / \Psi_{t-1}) = \hat{\xi}_{t/t-1}.$$

Consider that we want to forecast the value of y_t : $\hat{y}_{t/t-1} = \hat{E}(y_t / x_t, \Psi_{t-1})$

From (3.5.1.2) we notice that $\hat{E}(y_t) = A'x_t + H'\xi_t$. From the law of iterated projections and using the previous expressions we conclude that

$$\hat{y}_{t/t-1} = A'x_t + H'E(\xi_t / x_t, \Psi_{t-1}) = A'x_t + H'\hat{\xi}_{t/t-1} \quad (3.5.1.15)$$

The error of this forecast can be derived analytically with the help of (3.5.1.2) equation. More specifically,

$y_t - \hat{y}_{t/t-1} = A'x_t + H'\xi_t + w_t - A'x_t - H'\hat{\xi}_{t/t-1} = H'(\xi_t - \hat{\xi}_{t/t-1}) + w_t$ with mean square error given by the following formula:

$$E\left[(y_t - \hat{y}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right] = E\left[H'(\xi_t - \hat{\xi}_{t/t-1})(\xi_t - \hat{\xi}_{t/t-1})'H\right] + E\left[w_t w_t'\right] \quad (3.5.1.16).$$

Cross product terms have disappeared, since $E\left[w_t(\xi_t - \hat{\xi}_{t/t-1})'\right] = 0$ because we have assumed that w_t is uncorrelated with ξ_t. Furthermore, since $\hat{\xi}_{t/t-1}$ is a linear

function of Ψ_{t-1} it must be also uncorrelated with w_t. Using (3.5.1.4) and (3.5.1.13), equation (3.5.1.16) can be written as follows:

$$E\left[(y_t - \hat{y}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right] = H'P_{t/t-1}H + R \quad (3.5.1.17)$$

Updating The Inference About ξ_t

Next the inference about the current ξ_t is updated on the basis of the observation of y_t to produce $\hat{\xi}_{t/t} = \hat{E}(\xi_t / y_t, x_t, \Psi_{t-1}) = \hat{E}(\xi_t / \Psi_t)$. This can be calculated by using the formula for updating a linear projection:

$$\hat{\xi}_{t/t} = \hat{\xi}_{t/t-1} + \left\langle E\left[(\xi_t - \hat{\xi}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right]\right\rangle$$

$$\times \left\langle E\left[(y_t - \hat{y}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right]\right\rangle^{-1} \times (y_t - \hat{y}_{t/t-1}) \ (3.5.1.18).$$

But

$$E\left\langle(\xi_t - \hat{\xi}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right\rangle = E\left\langle\left[\xi_t - \hat{\xi}_{t/t-1}\right]\left[H'((\xi_t - \hat{\xi}_{t/t-1}) + w_t\right]'\right\rangle =$$

$$= E\left[(\xi_t - \hat{\xi}_{t/t-1})(\xi_t - \hat{\xi}_{t/t-1})'H\right] = P_{t/t-1}H \ (3.5.1.19)$$

By appropriate substitutions we end up to the following formula for updating inference about ξ_t:

$$\hat{\xi}_{t/t} = \hat{\xi}_{t/t-1} + P_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}.(y_t - A'x_t - H'\hat{\xi}_{t/t-1}). \quad (3.5.1.20)$$

The mean square error associated with the updated projection, which is denoted by $P_{t/t}$, can be found as follows:

$$P_{t/t} \equiv E\left[(\xi_t - \hat{\xi}_{t/t})(\xi_t - \hat{\xi}_{t/t})'\right] = E\left[(\xi_t - \hat{\xi}_{t/t-1})(\xi_t - \hat{\xi}_{t/t-1})'\right] - \left\langle E\left[(\xi_t - \hat{\xi}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right]\right\rangle$$

$$\times \left\langle E\left[(y_t - \hat{y}_{t/t-1})(y_t - \hat{y}_{t/t-1})'\right]\right\rangle^{-1} \times \left\langle E\left[(y_t - \hat{y}_{t/t-1})(\xi_t - \hat{\xi}_{t/t-1})'\right]\right\rangle^{-1} =$$

$$= P_{t/t-1} - P_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}H'P_{t/t-1} \quad (3.5.1.21)$$

Producing a Forecast of ξ_{t+1}

Next, the state equation (3.5.1.1) is used to forecast ξ_{t+1} :

$$\hat{\xi}_{t+1/t} = \hat{E}(\xi_{t+1}/\Psi_t) = F\,\hat{E}(\xi_t/\Psi_t) + \hat{E}(v_{t+1}/\Psi_t) = F\hat{\xi}_{t/t} + 0 \quad (3.5.1.22)$$

Substituting (3.4.20) into (3.4.22) we derive the following expression:

$$\hat{\xi}_{t+1/t} = F\hat{\xi}_{t/t-1} + FP_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t/t-1}) \quad (3.5.1.23).$$

The last expression of the (3.5.1.23) equation is called the measurement innovation, or the residual. The residual reflects the discrepancy between the predicted measurement and the actual measurement y_t. The matrix given by $FP_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}$ is known as the gain or blending factor denoted by K_t and is chosen in such a manner in order to minimize the a-posteriori error covariance of the state forecast. Consequently equation (3.5.1.23) becomes

$$\hat{\xi}_{t+1/t} = F\hat{\xi}_{t/t-1} + K_t(y_t - A'x_t - H'\hat{\xi}_{t/t-1}) \quad (3.5.1.24).$$

The mean square error of this forecast can be found from (3.5.1.22) and the state equation (3.5.1.1). Analytically the mean square error can be derived from the following formula:

$$P_{t+1/t} = E\left[\left(\xi_{t+1} - \hat{\xi}_{t+1/t}\right)\left(\xi_{t+1} - \hat{\xi}_{t+1/t}\right)'\right] = E\left[\left(F\xi_t + v_{t+1} - F\hat{\xi}_{t/t}\right)\left(F\xi_t + v_{t+1} - F\hat{\xi}_{t/t}\right)'\right] =$$

$$= FE\left[\left(\xi_{t+1} - \hat{\xi}_{t/t}\right)\left(\xi_{t+1} - \hat{\xi}_{t/t}\right)'\right]F' + E[v_{t+1}v_{t+1}'] = FP_{t/t}F' + Q \quad (3.5.1.25)$$

The cross products are again zero because of the initial assumptions. Furthermore substituting (3.5.1.21) into (3.5.1.25) we derive the following analytical form for the mean square error of the state forecast.

$$P_{t+1/t} = F\left[P_{t/t\text{-}1} - P_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}H'P_{t/t-1}\right]F' + Q. \,(3.5.1.26)$$

If we observe carefully the gain factor K_t we will see that as the observation error covariance approaches zero, the gain K_t weights the residual less heavily. Another way of thinking about the gain factor K_t is that as the measurement error covariance R approaches zero, the actual measurement is trusted more and more, while the predicted measurement is trusted less and less.

Summary of The Kalman-Filter Process and Remarks

To summarize, the Kalman-Filter starts with the unconditional mean and variance of ξ_1: $\hat{\xi}_{1/0} = E(\xi_1)$ and $P_{1/0} = E\left\langle \left[\xi_1 - E(\xi_1)\right]\left[\xi_1 - E(\xi_1)\right]'\right\rangle$. Typically these are given by $\hat{\xi}_{1/0} = 0$ and $vec(P_{1/0}) = \left[(I_{r^2} - (F \otimes F)\right]^{-1} veq(Q)$.

Then, we iterate on $\hat{\xi}_{t+1/t} = F\hat{\xi}_{t/t-1} + FP_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t/t-1})$ and $P_{t+1/t} = F\left[P_{t/t\text{-}1} - P_{t/t-1}H(H'P_{t/t-1}H + R)^{-1}H'P_{t/t-1}\right]F' + Q$ for $t = 1,2,...,T$. The value of $\hat{\xi}_{t+1/t}$ denotes the best forecast of ξ_{t+1} based on a constant and a linear function of $(y_t, y_{t-1},..., y_1, x_t, x_{t-1},..., x_1)$. The matrix $P_{t+1/t}$ gives the mean square error of this forecast. The forecast of y_{t+1} is given by $\hat{y}_{t+1/t} = A'x_{t+1} + H'\hat{\xi}_{t+1/t}$ with associated mean square error $E\left[(y_{t+1} - \hat{y}_{t+1/t})(y_{t+1} - \hat{y}_{t+1/t})'\right] = H'P_{t+1/t}H + R$.

The Kalman-Filter process estimates a process by using a form of feedback control: the filter estimates the state process at some time and then obtains feedback in the form of (noisy) measurements. As such, the Kalman-Filter equations fall into two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward in time the current state and the error covariance estimates in order to obtain a-priori estimates for the next time step. The measurement update equations are responsible for the feedback i.e. for incorporating a new measurement into the a-priori estimate to obtain an improved a-posteriori estimate. The time update equations can also be thought as predictor equations, while

the measurement update equations can be thought as corrector equations. Indeed the final estimation algorithm resembles that of a predictor-corrector algorithm for solving numerical problems. Schematically we can describe the Kalman-Filter algorithm as follows:
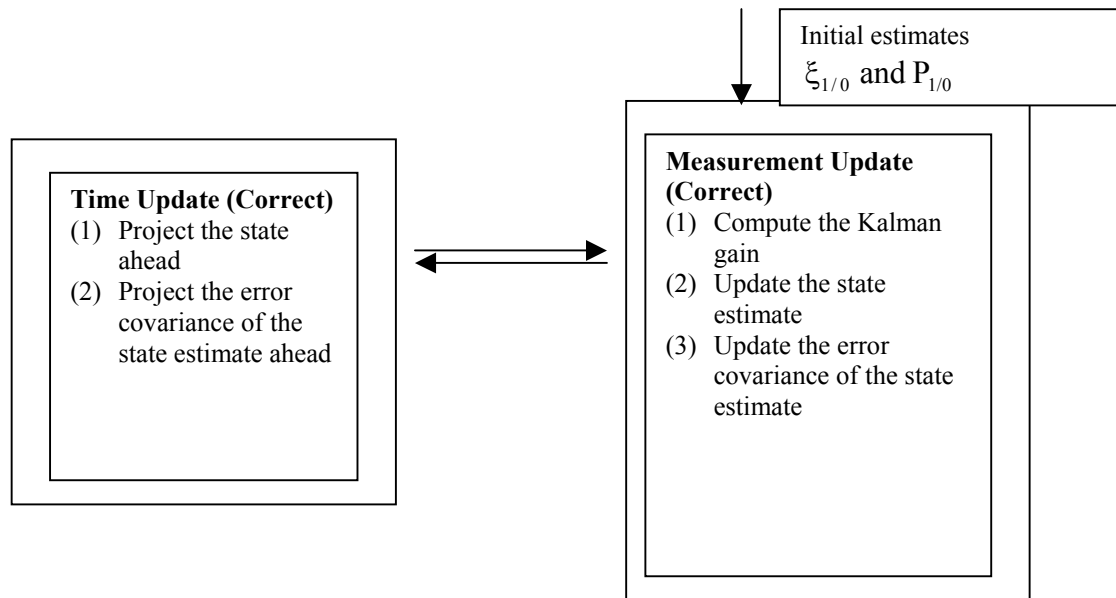

Figure 3.1: Graphical representation of the Kalman-filter algorithm.

Using the Kalman-Filter to Evaluate The Likelihood Function

The Kalman-Filter was previously motivated in terms of linear projections. The forecasts $\hat{\xi}_{t/t-1}$ and $\hat{y}_{t/t-1}$ are thus optimal within the set of forecasts that are linear in (x_t, Ψ_{t-1}) where $\Psi_{t-1} \equiv (y_{t-1}', y_{t-2}', ..., y_1', x_{t-1}', x_{t-2}', ..., x_1')'$. If the initial state ξ_1 and the innovations $(w_t, v_t)_{t=1}^{T}$ are multivariate Gaussian, then we can make the stronger claim that the forecasts $\hat{\xi}_{t/t-1}$ and $\hat{y}_{t/t-1}$ calculated by the Kalman-Filter are optimal among any functions of (x_t, Ψ_{t-1}). Moreover, if ξ_1 and $(w_t, v_t)_{t=1}^{T}$ are Gaussian, then the distribution of y_t conditional on (x_t, Ψ_{t-1}) is Gaussian with mean and variance given by the following expression:

$$y_t / x_{t,}, \Psi_{t-1} \sim N\left(\left(A'x_t + H'\hat{\xi}_{t/t-1}\right), \left(H'P_{t/t-1}H + R\right)\right); \text{ that is}$$

$$f_{Y_t / X_t, \Psi_{t-1}} (y_t / x_t, \Psi_{t-1}) = (2\pi)^{-n/2} \left| H'P_{t/t-1}H + R \right|^{-1/2} \tag{3.5.1.27}$$

$$\times \exp \left\langle -\frac{1}{2} \left(y_t - A'x_t - H'\hat{\xi}_{t/t-1} \right)' \left(H'P_{t/t-1}H + R \right)^{-1} \left(y_t - A'x_t - H'\hat{\xi}_{t/t-1} \right) \right\rangle \text{ for } t = 1,2,...T$$

From (3.4.26), we can construct the log-likelihood that has the following form:

$$\sum_{t=1}^{T} \log f_{Y_t} /_{X_t, \Psi_{t-1}} \left(y_t / x_t, \Psi_{t-1} \right). \tag{3.5.1.28}$$

Expression (3.5.1.27) can be maximized numerically with respect to the unknown parameters in the matrices F, Q, A, H and R.

Smoothing

The Kalman-Filter was previously motivated as an algorithm for calculating a forecast of the state vector ξ_t as a linear function of previous observations $\hat{\xi}_{t/t-1} \equiv \hat{E}(\xi_t / \Psi_{t-1})$. The matrix $P_{t/t-1}$ represents the mean square error of the forecast and can be estimated by using the following formula:

$$P_{t/t-1} \equiv E \left[\left(\xi_t - \hat{\xi}_{t/t-1} \right) \left(\xi_t - \hat{\xi}_{t/t-1} \right)' \right].$$

A goal might then be to form an inference about the value of ξ_t based on the full set of data collected, including observations on $y_t, y_{t+1},..., y_T, x_t, x_{t+1},..., x_T$. Such an inference is called the smoothed estimate of ξ_t, denoted by $\hat{\xi}_{t/T} \equiv \hat{E}(\xi_t / \Psi_T)$. For example, data on GNP[15] from 1954 through 1990 might be used to estimate the value that ξ took on in 1960. The mean square error of this smoothed estimate is denoted

with $P_{t/T} \equiv E \left[\left(\xi_t - \hat{\xi}_{t/T} \right) \left(\xi_t - \hat{\xi}_{t/T} \right)' \right]$ (3.5.1.29).

Consider the estimate of ξ_t based on observations through date t, $\hat{\xi}_{t/t}$. Suppose we were subsequently told the true value of ξ_{t+1}. From the formula for updating a linear projection, the new estimate of ξ_t could be expressed as:

---

[15] Gross National Product.

$$\hat{E}(\xi_t / \xi_{t+1}, \Psi_t) = \hat{\xi}_{t/t} + \left\langle E\left[(\xi_t - \hat{\xi}_{t/t})(\xi_{t+1} - \hat{\xi}_{t+1/t})' \right] \right\rangle$$

$$\times E\left\langle \left[(\xi_{t+1} - \hat{\xi}_{t+1/t})(\xi_{t+1} - \hat{\xi}_{t+1/t})' \right] \right\rangle^{-1} \times \left(\xi_{t+1} - \hat{\xi}_{t+1/t} \right) (3.5.1.30)$$

The first term in the product on the right hand side of (3.5.1.30) can be written as

$$E\left[(\xi_t - \hat{\xi}_{t/t})(\xi_{t+1} - \hat{\xi}_{t+1/t})' \right] = E\left[(\xi_t - \hat{\xi}_{t/t})(F\xi_t + v_{t+1} - F\hat{\xi}_{t/t}) \right].$$

Furthermore, v_{t+1} is uncorrelated with ξ_t and $\xi_{t/t}$. Thus,

$$E\left[.(\xi_t - \hat{\xi}_{t/t})(\xi_{t+1} - \hat{\xi}_{t+1/t})' \right] = E\left[(\xi_t - \hat{\xi}_{t/t})(\xi_t - \hat{\xi}_{t/t})' F' \right]. = P_{t/t}F' \, (3.5.1.31).$$

Substituting (3.5.1.31) and the definition of $P_{t+1/t}$ into (3.5.1.30) we obtain the following expression:

$$\hat{E}(\xi_t / \xi_{t+1}, \Psi_t) = \hat{\xi}_{t/t} + P_{t/t}F'P_{t+1/t}^{-1}(\xi_{t+1} - \hat{\xi}_{t+1/t})$$

Defining $J_t \equiv P_{t/t}F'P_{t+1/t}^{-1}$ (3.5.1.32)

we have that $\hat{E}(\xi_t / \xi_{t+1}, \Psi_t) = \hat{\xi}_{t/t} + J_t(\xi_{t+1} - \hat{\xi}_{t+1/t})$

The linear projection (3.5.1.32) turns out to be the same as $\hat{E}(\xi_t / \xi_{t+1}\Psi_T) \, (3.5.1.33)$. That means that knowledge of y_{t+j} or x_{t+j} would be of no added value if we already knew the value of ξ_{t+1}.

It follows from the law of iterated projections that the smoothed estimate, $\hat{E}(\xi_t / \Psi_T)$, can be obtained by projecting (3.5.1.33) on Ψ_T. In calculating this projection, we need to think carefully about the nature of the magnitudes in (3.5.1.33). The first term, $\hat{\xi}_{t/t}$, indicates a particular exact linear function of Ψ_T; the coefficients of this function are constructed from population moments and these coefficients should be viewed as deterministic constants. The projection of $\hat{\xi}_{t/t}$ on Ψ_T is thus still $\hat{\xi}_{t/t}$. The term J_t is also a function of population moments, and so is again treated as

deterministic quantity. The term $\hat{\xi}_{t+1/t}$ is another exact linear function of Ψ_T. Thus the projection on Ψ_T is given by the following expression:

$$\hat{E}(\xi_t / \Psi_T) = \hat{\xi}_{t/t} + J_t \left[E(\xi_{t+1} / \Psi_T) - \hat{\xi}_{t+1} \right] \text{ or}$$

$$\hat{\xi}_{t/T} = \hat{\xi}_{t/t} + J_t (\hat{\xi}_{t+1/T} \, \hat{\xi}_{t+1/t}) \qquad (3.5.1.34)$$

Thus, the sequence of smoothed estimates $\left\langle \hat{\xi}_{t/T} \right\rangle_{t=1}^{T}$ is calculated as follows:

First the Kalman-Filter is calculated and the sequences $\left\langle \hat{\xi}_{t/t} \right\rangle_{t=1}^{T}$ $\left\langle \hat{\xi}_{t+1/t} \right\rangle_{t=0}^{T-1}$ $\left\langle P_{t/t} \right\rangle_{t=1}^{T}$

$\left\langle P_{t+1/t} \right\rangle_{t=0}^{T-1}$ are stored. The smoothed estimate for the final date in the sample, $\left\langle \hat{\xi}_{T/T} \right\rangle$

is just the last entry $\left\langle \hat{\xi}_{t/t} \right\rangle_{t=1}^{T}$. Next we generate $\left\langle J_t \right\rangle_{t=1}^{T-1}$. Then we use (3.5.1.34) for

$t = T - 1$ to calculate $\hat{\xi}_{T-1/T} = \hat{\xi}_{T-1/T-1} + J_{T-1}(\hat{\xi}_{T/T} \, \hat{\xi}_{T/T-1})$. Now that

$\hat{\xi}_{T-1/T}$ has been calculated we use again (3.5.1.34) for $t = T - 2$. Proceeding backward through the sample in this fashion permits calculation of the full set of smoothed estimates.

As far as it concerns the mean square error associated with the smoothed estimate we obtain the following:

$$\xi_t - \hat{\xi}_{t/T} = \xi_t - \hat{\xi}_{t/t} - J_t \hat{\xi}_{t+1/T} + J_t \hat{\xi}_{t+1/t}$$
$$\text{or}$$
$$\xi_t - \hat{\xi}_{t/T} + - J_t \hat{\xi}_{t+1/T} = \xi_t - \hat{\xi}_{t/t} + J_t \hat{\xi}_{t+1/t}$$

Multiplying the previous equation by its transpose and taking expectations, we obtain

$$E\left[(\xi_t - \hat{\xi}_{t/T})(\xi_t - \hat{\xi}_{t/T})' \right] + J_t E\left[(\hat{\xi}_{t+1/T} \, \hat{\xi}_{t+1/T}') \right] J_t' = \qquad (3.5.1.35)$$

$$= E\left[(\xi_t - \hat{\xi}_{t/t})(\xi_t - \hat{\xi}_{t/t})' \right] + J_t E\left[(\hat{\xi}_{t+1/t} \, \hat{\xi}_{t+1/t}') \right] J_t'$$

The cross-product terms have been disappeared from the left side because $\hat{\xi}_{t+1/T}$ is a

linear function of Ψ_T and so is uncorrelated with the projection error $\xi_t - \hat{\xi}_{t/t}$.

Similarly, on the right hand side, $\hat{\xi}_{t+1/t}$ is uncorrelated with $\xi_t - \hat{\xi}_{t/t}$. Consequently equation (3.5.1.35) states that:

$$P_{t/T} = P_{t/t} + J_t \left\langle -E\left[(\hat{\xi}_{t+1/T}\,\hat{\xi}_{t+1/T}{}')\right] + E\left[(\hat{\xi}_{t+1/t}\,\hat{\xi}_{t+1/t}{}')\right]\right\rangle J_t{}'. \quad (3.5.1.36)$$

With the help of (3.5.1.36) we obtain the mean square error for the smoothed estimate which is given by the following formula:

$$P_{t/T} = P_{t/t} + J_t (P_{t+1/T} - P_{t+1/t}) J_t{}' \quad (3.5.1.37).$$

Again, this sequence is generated by moving through the sample backward, starting with $t = T - 1$.

The French Model

The principle underlying the retropolation in the French case is as follows. We are trying to work out for a period of T years the measurement of an item that is known for only k years (k<T) in two accounting systems simultaneously and for T years in one of the systems. Schematically the previous case can be described by the following figure.

T years unknown under new rules   k years known under new rules

T+k years known under old rules
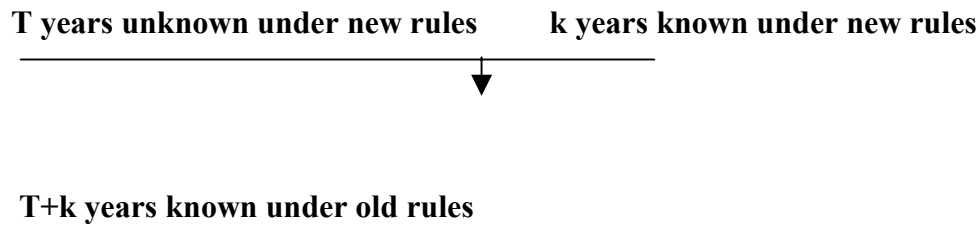
For this purpose France and more specifically the Institut National de la Statistique et de Etudes Economiques (INSEE) proposes to use a linear model linking the variables in the two accounting systems and from this to calculate the conditional (or linear) expectation of the non-observed variables as a function of the set of values available under the former system.

Let x_t be a vector of the observed values available at date t of N accounting magnitudes obtained in the former base. Assume now that a new measure X_t is introduced. Our objective is to construct the time-series for the past according to the

new measure **X**. In other words what we are looking for, is to calculate the following conditional expectation.

$$E(X_t / x_1, x_2, ..., x_T, x_{T+1}, ...x_{T+k}, X_{T+1}, ...X_{T+k})$$

Experience and several trials have led France to adopt two types of short-term annual dynamic linear relationship between x_t and X_t having the following form:

$$
\begin{cases}
\Delta X_t = \alpha \Delta x_t + \beta \Delta x_{t-1} + \chi \Delta X_{t-1} + \varepsilon_t \\
\Delta x_t = \delta \Delta x_{t-1} + \phi(\Delta X_{t-1}) + \eta_t \\
\text{or} \\
X_t - \alpha x_t = +\beta \Delta x_t + \gamma \Delta x_{t-1} + \chi(X_{t-1} - \alpha x_{t-1}) + \varepsilon_t \\
\Delta x_t = \delta \Delta x_{t-1} + \phi(X_{t-1} - \alpha x_{t-1}) + \eta_t
\end{cases}
$$

$$\text{with } \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma & 0 \\ 0 & \Sigma \end{pmatrix} \right).$$

The model above contains non-observable variables and the estimation of the parameters is carried out using the Kalman-Filter method. However, this type of estimation remains problematical from a numerical standpoint since the likelihood may show local maximum.

The estimation procedure of the French retropolation model can be described by the following algorithm:

$$
\begin{aligned}
X_t = \alpha x_{t-1} + \beta X_{t-1} + \varepsilon_t \\
x_t = \delta x_{t-1} + \phi X_{t-1} + \eta_t
\end{aligned}
\text{ with } \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma & \vartheta \\ \vartheta & \sigma \end{pmatrix} \right).
$$

Expressing the model in state-space form-necessary in order to apply the Kalman-Filter method- we derive the following model:

$$
\begin{aligned}
z_t &= \Phi z_{t-1} + v_t \\
y_t &= (0, I_N)z_t, K < t \le T + K \\
y_t &= z_t, 1 < t \le K
\end{aligned}
\text{ where } z_t = \begin{pmatrix} X_t \\ x_t \end{pmatrix}, \Phi = \begin{pmatrix} \alpha & \beta \\ \delta & \phi \end{pmatrix} \text{ and } v_t = \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}.
$$

We want to express the likelihood of $(y_2, y_3, ..., y_{T+K-1}, y_{T+K})$. Applying the properties of conditional expectation, we obtain:

$$\ln(l(\Phi,\Sigma/y_1,...,y_{T+K})) = \ln(f(y_2/y_1,\Phi,\Sigma))\ln(f(y_3/y_1,y_2,\Phi,\Sigma))...$$
$$\ln(f(y_{T+K}/y_{T+K-1},...,y_1\Phi,\Sigma))$$
$$\ln(l(\Phi,\Sigma/y_1,...,y_{T+K})) = -\frac{TN+(k-1)(N+1)}{2}\ln(2\pi) - \frac{k-1}{2}\ln(\det\Sigma)$$
$$-\frac{1}{2}\sum_{K+1}^{T+K}\ln(\det\Sigma(t/t-1)) - \frac{1}{2}\sum_2^K(y_t - \Phi y_{t-1})'\Sigma^{-1}(y_t - \Phi y_{t-1}) -$$
$$\frac{1}{2}\sum_{K+1}^{T+K}(y_t - y_{t/t-1})'\Sigma(t/t-1)^{-1}(y_t - y_{t/t-1})$$

where y_t represents the conditional likelihood of y_t for the entire set of observations to date s and $\Sigma(t/s)$ is the conditional variance for the same set of information. These quantities for $(t > K)$ are calculated by the following steps.

- Initialization
$$V(z_{k/k-1}) = \Sigma$$
$$z_{k/k-1} = y_k$$

$$z_{t/t-1} = \Phi z_{t-1/t-1}$$
$$V(z_{t/t-1}) = \Phi V(z_{t-1/t-1})\Phi' + \Sigma$$

- Conditional Expectation $\quad y_{t/t-1} = (0, I_N)z_{t/t-1}$

$$V(y_{t/t-1}) = (0, I_N)V(z_{t/t-1})\begin{pmatrix} 0 \\ I_N \end{pmatrix}$$

$$z_{t/t} = z_{t/t-1} + V(z_{t/t-1})\begin{pmatrix} 0 \\ I_N \end{pmatrix}V(y_{t/t-1})^{-1}(y_t - y_{t/t-1})$$

- Correction $\quad V(z_{t/t}) = V(z_{t/t-1}) - V(z_{t/t-1})\begin{pmatrix} 0 \\ I_N \end{pmatrix}V(y_{t/t-1})^{-1}\begin{pmatrix} 0 & I_N \end{pmatrix}V(z_{t/t-1})$.

Once the parameters have been estimated, it is possible to construct by a smoothing process the likelihood of $(1,0_N)z_t$ conditional on the whole of the available observations which is what we were looking for. This is obtained by the following procedure:

Smoothing: $z_{t/T+K} = z_{t/t} + V(z_{t/t})\Phi V(z_{t+1/t})^{-1}(z_{t+1/T+K} - z_{t+1/t})$.

Software for The French Retropolation Method

A retropolation program was developed to provide automated estimates of time series. France's retropolation software is supported by SAS software and IML[16] and it is developed for MVS operating system.

It has an understandable manual that provides definitions of the basic concepts as well as examples.

There is a clear way of inputting the data consisting of the series from the old base as well as the series from the new base.

The software provides a number of statistics for the evaluation of the candidate models. Statistics of this type are the standard deviations of the parameter estimations (the smaller it is the better), the standard deviation of the period of analysis (the smaller it is the smaller the degree of imprecision), the Bayes information Criterion for assessing how good is the fit of the estimated models as well as a number of other criteria for evaluating the convergence of the algorithm. Moreover the software produces graphs of the retropolated series and the old base series which provide an additional help in the evaluation of the candidate models. Although there is an on line help and a friendly to the user environment the present retropolation program requires a familiarity with the advanced statistical programs such as SAS and S+ and also knowledge on basic statistical concepts. As a result there is a need of training in order the user to be able to use the macros and to program in a satisfactory level.

The big advantage of this software is that by using it we are in the position to evaluate all candidate models and decide which is the optimal one. However, we must be very careful when speaking for maximization of the likelihood or for convergence because instead of a global maximum we might find a local one. Also the evaluation of the candidate models is based on some statistical criteria but these are not the only ones. The best solution in this case is to adopt a model choice approach i.e. run different models and decide which is the best.

---

[16] Interactive Matrix Language.

3.6 Concluding Remarks

Eurostat is now revising the backward calculation methods in order to be able to make proposals to other Member States to develop a relevant methodology. These proposals aim at the better harmonization of the backward calculation techniques and consequently to the better over space comparability of the results

Attempting to make some concluding remarks about the retropolation methods in France and Netherlands we can say the following:

The Netherlands approach is based on the benchmark years/interpolation method. In this method, after specifying the corrections for the benchmark years, interpolation is applied in order to specify the corrections for the intermediate years.
The French approach is different in that the estimation of the intermediate years is based on a linear model that links the variable in the new accounting system and the variables of the accounting system before revision. The estimates of the linear model are obtained via the Kalman–Filter algorithm.

The best solution for the backward calculation is the full annual backward calculation. However, this technique requires a very good system of basic statistics, very good and detailed knowledge of the national accounting system and is very time and staff consuming. Under these restrictions benchmark years/interpolation method seems to offer a good alternative solution. Although benchmark years method is a more rough and mechanical method, it can give good solutions to the problem of backward calculation but only under certain quality requirements.