A Unified Euro Area Database
(UEADA)

The Monthly Dataset, 1987-2001

In contrast to the United States, where analysts can easily access well-established and large databases, no such databases yet exists for Europe. We had therefore to consult and evaluate many different sources: among others, national statistical institutes, the OECD and the Eurostat statistics; from these we collected and examined a large number of series, organizing them in a detailed dataset.

The final database whose richness of properly organized and monthly updated information could makes it a particularly useful tool for further research has been organized into the following eleven homogeneous blocks, corresponding to different major sectors:

- industrial production
- producer prices
- consumer prices
- monetary aggregates
- interest rates
- financial variables
- exchange rates
- surveys by the European Commission
- surveys by national institutes
- external trade
- labour markets

On the whole, they provide a comprehensive description of the European economy. Each block contains time series for Germany, France, Italy, Spain, The Netherlands, Belgium and when available, for the euro area as a whole. Indeed, since a business cycle indicator for the euro area reflects economy wide fluctuations common across countries, one should collect data covering a wide variety of sectors for all European economies. Unfortunately, data limitations forced us to restrict the focus on the six largest countries that, nonetheless, accounted for more than 90% of the euro aggregate GDP in 2000. Some key macroeconomic series not directly referred to the euro area were also included to capture phenomena that might be relevant to explain fluctuations across Europe; examples include are oil and raw material prices and some indicators of the business cycle in other large economies (UK, US and Japan).

Altogether the database contains about 3000 time series. Among these we selected only those variables that satisfy two crucial requirements: one concerning the length of the series and the other their homogeneity over time and across countries. As regards the first one, the largest common sample for the dataset spans the period January 1987 to March 2001.

Although many time series are available for a longer period, the decision to set the starting date in 1987 is the result of a trade-off between obtaining richer time series information and maintaining a large cross-sectional dimension for the dataset.

As for the second requirement (i.e. homogeneity over time and across countries) we selected variables from each of the eleven blocks, trying to maintain, for each of them and wherever possible, a common breakdown for all countries. In some cases to obtain series of sufficient length we had to splice together series covering shorter time spans (for example HICP and CPI or Pan-German with West German data), trying to match definitions and levels of aggregation as closely as possible. Whenever we had to intervene with some kind of manipulation to obtain time series of the desired quality, the strategy adopted for data reconstruction was the following. For the most recent years series were collected from Eurostat or the ECB, since these institutions coordinate national sources in the process of statistical harmonization. Then other international institutions (like the OECD) or national sources (e.g. INSEE) were used to obtain series of sufficient length or to cover important economic phenomena with the desired detail.

As a last step, in order to avoid overweighting a single country or a particular economic sector, in selecting time series we tried to maintain a satisfactory balance in terms of the number of series across countries and blocks. Nevertheless, closely pursuing this criterion with the available statistics would have forced us to work with a minimal common set of indicators, thereby forgoing much information. This is the reason why, to meet the requirement of a large database, we preferred to relax the condition on perfect balancing.

Data Sources

This describes the main guidelines followed setting up the database and, in particular, each of the blocks into which it has been split. The general strategy adopted was to collect data for most recent years from Eurostat and the European Commission, whenever they were available: these sources should grant a proper statistical harmonization across countries for the information released. Nevertheless, many other international sources and national institutions were consulted in order to construct a dataset that gives a comprehensive account of the economic phenomena for the largest European countries. In these cases attention was paid to gather data of homogeneous quality. Finally the database has been organized in a way that allows monthly updates of all the time series therein: this is obviously a fundamental requirement in view of monthly releases of the cyclical indicators built upon it.

The trading days and seasonally adjusted series on Industrial Production were extracted from the Eurostat database, organized according to the NACE Rev. 1 classification method and generally covering a sufficiently long time span. Nevertheless in some cases earlier data were collected from the CECO database, according to the SIC classification; the Eurostat time series were then linked backward trying to match definitions and disaggregations as closely as possible. In spite of this, most industrial production time series for The Netherlands and for Belgium start only in the nineties and therefore cannot be used to perform the dynamic factor model estimation.

For producer prices we replicated the sectoral breakdown used for industrial production (NACE Rev.1); in doing so we resorted to the Eurostat database on PPIs and on some national sources, such as ISTAT for Italy and INSEE for France. Consumer price series are the result of a link between the most recent HICP data available from Eurostat, starting in 1995, and a combination of earlier data from either the main economic indicators database of the OECD, or national statistical institutes (ISTAT, INSEE) and Datastream.

The monetary block includes various definitions of money aggregates (M1, M2 and M3) for the largest European economies; besides this, a wide variety of interest rates was gathered covering both short and long term government bonds, bank deposits and bank loans, When available, some spreads between interest rates were included too, especially for the Italian economy. Effective exchange rates were also collected for all of the countries considered, both in real and in nominal terms. The main sources consulted for the variables belonging to this block are the BIS (Bank of International Settlements), the ESCIB and some national institutions.

Harmonizing the data collected by national sources, the European Commission monthly provides seasonally adjusted business and households survey results, both for the Euro Area and for each member country. Constructions, retail trade and manufacturing sectors are investigated and the economic sentiment indicator is obtained to synthesise the overall business climate. Time series reporting balances of the answers start in the mid eighties and are regularly updated, some of them regard questions addressed quarterly to economic agents and are therefore not exploited in the present work. National institutions (e.g. IFO for Germany, INSEE for France, ISAE for Italy etc.) survey datasets cover longer time spans and a deeper level of disaggregation of economic activities; for these reasons they were included in our database too, in addition to those provided by the European Commission.

Relevant business cycle information can be extracted from data that are not classifiable among the previously described sets of time series. A further group was consequently formed, containing a miscellaneous collection of variables concerning many different economic phenomena, such as passenger car and other vehicles registrations, new companies formation, declarations of bankruptcy, share-price indexes, orders, turnovers, construction permits, rail transportations of passengers and goods and many others. Due to the particular nature of these variables, it was not always possible to collect them for each country; as a consequence, this set of series is not perfectly balanced but, nonetheless, proved to be useful.

It has been particularly difficult to obtain labour-market variables satisfying the requirements listed in section 3.1 and needed for the estimation of the model. OECD and BIS databases were consulted, obtaining sufficiently exhaustive information concerning the unemployment in all European countries. Although with less detail, time series on wages and unit labour costs were found, whilst very little information about vacancies is available.

Finally, exports and imports time series - especially regarding consumer, intermediate and capital goods - were extracted from BIS and OECD datasets to constitute the Trade block. More interestingly this block of data includes time series on the trade volume between each euro area countries and its main commercial partners.

The dynamic behaviour of the series collected is remarkably non-homogeneous. Most of them are raw series, others have been adjusted to take into account working days effects and, finally, a few are available only in a seasonally adjusted version. Preliminary inspection reveals that our series are not affected by the same kind of non-stationarity. Given the large number of series in the panel careful individual treatment of non-stationarity was not feasible. Rather, we followed an automatic procedure treating in the same way all the series of a given economic class (e.g. industrial production, consumer prices and so on). Then we checked whether this resulted in an improper treatment of the data, such as over-differencing, incomplete removal of outliers or inadequate seasonal adjustment. When this was the case, if the problem could not be fixed with some ad hoc adjustment, the variable was discarded from the dataset.

Our data treatment procedure can be detailed in four steps as follows:

1. We detected and removed outliers from each series using Tramo, a procedure developed by Gomez and Maravall; in particular we focused on transitory changes, level shifts and additive outliers. The same procedure permitted adjustment for working days effects, whenever requested. Once these deterministic factors were removed, each series was seasonally adjusted using seasonal dummy variables. To take into account the possibility of changes in the seasonal pattern over time, the dummies were also coupled with a linear time trend. We did not resort to other more sophisticated procedures (e.g. Seats or X12) to avoid the use of bilateral filters, which would imply large revisions in the seasonally adjusted series and therefore in the cyclical indicator.[16]

2. The adjusted data were further inspected to make sure that the procedure described above successfully removed all major irregularities. In a few cases we had to drop time series that even after the first step displayed major breaks or other inconsistencies that could not be accounted for and that were therefore attributed to the poor quality of the data.

3. In order to estimate the cross spectral density matrix the series need to be covariance stationary. The transformation inducing stationarity was applied to each outliers free and seasonally adjusted series. The first log difference was taken for industrial productions, trade variables, financial series, monetary aggregates and labour market variables; exceptions were made for some series (e.g. unemployment rates) for which first differences of natural

values were taken. First difference of natural values was also applied to business and household survey responses and to the vast majority of interest rates. Real interest rates and the spreads between long and short-term nominal interest rates did not need any transformation.[17] The order of integration of price variables is controversial since the choice between I(1) and I(2) models is not always clear-cut. Given that in the majority of cases an I(1) classification seems appropriate, we decided in favour of considering all price indexes as I(1).