



STATISTICS DIRECTORATE

STD/STESEG(2003)8
For Official Use

OECD Short-Term Economic Statistics Expert Group

STESEG TASK FORCE ON TIMELINESS AND BENCHMARKING

USE OF MODELLING TECHNIQUES AND FLASH ESTIMATES AT THE FEDERAL STATISTICAL OFFICE OF GERMANY

Paris, 26-27 June 2003

*Prepared by: Kai Lorentz
Federal Statistical Office of Germany*

Submitted to the Expert Group under item 5 on the draft agenda.

Contact: Richard McKenzie (richard.mckenzie@oecd.org), Statistics Directorate, OECD

JT00145784

USE OF MODELLING TECHNIQUES AND FLASH ESTIMATES AT THE FEDERAL STATISTICAL OFFICE OF GERMANY

Introduction

In this short paper we like to present the methods which are currently in place at the Federal Statistical Office of Germany and also the methods which are planned to be introduced in the future. Within the work program of the STESEG task force on timeliness and benchmarking, this paper is a contribution to the task # 6 – the use of modelling techniques and flash estimates.

In general, German official statistics are almost always organised as a division of labour between the Federal Statistical Office of Germany (DESTATIS) and the Statistical Offices of the German Länder. Data collection and presentation of (partial) results of the Länder is in general task of the Länder whereas methodological issues (such as sampling plans, etc.) and presentation of results on national level are within the workload of DESTATIS. This division of labour between DESTATIS and the regional offices is in large part a consequence of general German law and takes into account the wide political and administrative independence of the German Länder. As a consequence, it is almost always the case that German surveys have to be designed in a way that they deliver results for the Länder up to a given or desired accuracy.

As far as timeliness is concerned, efforts have to be undertaken by both parts - DESTATIS as well as the Statistical Offices of the Länder. As a large part of the production process resides within the Statistical Offices of the Länder (from data collection over plausibility control up to regional aggregation), the participation of the Statistical Offices of the Länder is essential for success of such efforts.

From the point of view of the DESTATIS, it would be desirable to have more timely results on the national aggregation level. So for example the question has a long tradition in Germany, whether survey samples could be introduced which deliver more timely results on national aggregation level only, but with a considerable smaller sampling size than the actual surveys. For the above mentioned political reasons the efforts were not successful until now. So, another approach is widely followed at the moment in official German statistics: For the publication of first (estimated) results, only the figures of some of the German Länder are considered and the remaining are estimated. DESTATIS analyses which regional figures are the most significant for the national aggregation of the statistic in question and defines an estimation method - usually based on a regression- to be applied. So for each statistic, a formal agreement between DESTATIS and the Statistical Offices of the Länder in question is made to the extent that these regional offices make the effort to deliver their results more timely to DESTATIS. DESTATIS in turn aggregates these figures and calculates the estimation to be published. Afterwards, when the results from all of the 16 states are available, DESTATIS publishes the revised figures for the national aggregation level.

Domestic Product

Domestic Product and national account data is for sure among the most important economic indicators. In the past, quarterly national account data was published at T+55. With the beginning of this year (first quarter), the process was speeded up to deliver results at T+45.

Calculation of the German national account data requires the input of more than 80 individual statistics or indicators, some of them produced in house, others taken from external sources. To speed up the process, it is therefore necessary to receive more timely the results of all the statistics and indicators, which contribute to the domestic product. As these statistics are not all controlled by DESTATIS, other approaches were considered. So, there was initiated an investigation project about the possibility of introducing an ARIMA based nowcasting method for the estimation of the German Domestic Product, with the aim to deliver estimated figures at T+30.

In this project, each of the more than 80 contributing statistics is analysed separately by well known time series techniques. Seasonal adjustment is considered for most of the indicators and ARIMA models are estimated for the adjusted series. For the monthly time series involved, the possibility of estimating the third month of a quarter from the figures of the first and the second as well as on the past quarter figures, is investigated. For some time series with small variation, estimates are based only on annual change rates.

At the end, there results a variety of more than 80 different nowcasting methods for the contributing statistics. Quality measures such as the Mean Absolute Revision (MAR) and the Root Mean Square Error (RMSE) are considered to evaluate the nowcasting figures for past intervals.

It is yet not possible to report on final results of the project as investigations are still in progress. But it should be clear that the models involved for nowcasting will not remain stable over time. That is, model parameters will have to be recalculated periodically upon recent prolongation of the time series.

It is also still under consideration, whether the econometric model should be amplified, i.e. whether further economic indicators should be used for the calculation or not.

Prices

The German Consumer Price Index (CPI) is one of the most timely published Price Indices within Europe. The timeliness of this figure is of great importance for the users, especially for the European Central Bank and the German Bundesbank. The published figure is an estimation based on a regression model from partial results of the six most important (significant) German Länder (Nordrhein-Westfalen, Hessen, Baden-Württemberg, Bayern, Brandenburg and Sachsen).

Only the total index is estimated, no estimation is undertaken for the several sub indices which exist.

Data is collected by professional interviewers always around the 15th of each month from a sample. The sampling design is not a probabilistic but a typical one. Already by the end of the same month, the calculated index values of the above mentioned states are transmitted to DESTATIS. Here, the estimation of the national index is undertaken using a multi linear regression approach. In fact, three different multi linear regression estimates are made: one for the absolute value, one for the annual change rate and one for the monthly change rate. At the end, the three estimation results are adjusted to one final estimation result.

The regression coefficients are recalculated each month.

Experiences with this method had been very good within the past. The method turned out to be stable to produce an early estimate of the Consumer Price Index. In the last years, the revisions between the estimated and the final result was at a maximum of 0.1%. The mean absolute revision was at 0.03%.

Domestic Retail Trade Turnover

Recently, a regression based estimation method - similar to 3 - was introduced to deliver national results more timely.

Since 2002, estimated domestic retail trade turnover figures are published at T+30. As explained in the introduction, the estimate relies on partial preliminary results from some of the German Länder. With the statistical offices of these Länder, it was possible to come to arrangements that preliminary results of the Länder are passed timely DESTATIS.

The first estimation model in use is a simple linear regression model that tries to explain the national annual change rate from the annual change rates of the preliminary figures passed timely to DESTATIS. The regression parameters were estimated on the base of 33 monthly observations of the corresponding final results since preliminary figures were not available before the year of 2002.

Currently, refinements of the methodology are investigated, namely the use of a multi linear regression model and also the possibility to estimate totals instead of change rates. Final results are not yet available.

Revisions were published (mean revision and mean absolute revision) and indicated an acceptable quality of the estimated figures.

Hotel and Restaurant Industry Turnover

A similar model to 4 is likely to be introduced in the future.

The project did not really start yet, so we are not in a position to report on first results at the moment. The first step of the project will be to analyse the significance of the contributing results of the German Länder to the aggregated German national value. The analysis will be based on the ordinary figures of the Länder (not on early figures), which they passed to DESTATIS in the past. The idea is to find five to six Länder, which highly explain the German national result and thus allow the introduction of a regression model. Not only the significance is of importance, but also the stability within time to the explanation of the national value. I.e. the coverage of the totals of the five to six Länder have to be high.

After this, arrangements between DESTATIS and the Statistical Offices of the Länder will be made with the aim that the participating Länder will pass also early figures to DESTATIS.

Production Index

There was elaborated a conceptual study about the possibility of delivering more timely figures for the production index.

The study focussed on the introduction of an additional sample based on turnover. It was estimated, that about 2000 of the largest enterprises should be questioned in order to achieve the desired timeliness (around T+16, as in the US). However, accuracy was estimated not to be sufficiently high due to such a small sample size so that the results gained by such a sample survey were not considered reliable and high revision rates were expected, also because turnover and production are different economic variables.

Also, from the point of view of business cycle analysts, the production index in general is not considered to deliver early indications on economic turning points. The new orders index gives forecast information about the business cycle, whereas turnover is a lagging indicator. So the question arises whether early

estimates of the production index are really necessary and whether the additional costs are justified. At the end, it was decided to cancel the project.

Production in Construction

Preliminary investigations are undertaken to analyse the possibility of the introduction of a more timely estimation.

The approach followed is very similar to the one for the Consumer Price Index. As the project did not progress very far, we can not report on first results.

Foreign Trade

- a) Between 1998 and 1999 there had been realised a study by DESTATIS, financed by Eurostat, to deliver more timely monthly results for the intra European trade.

Such (aggregated) figures are currently available at T+40. The idea was to look at a (shorter) fixed time lag, say T+20, and to consider only the figures of those units, which already had entered the processing system until that date (e.g. through faster electronic transmission) and to estimate the total figures at a high aggregation level. This approach resulted in considering a non probabilistic sample. For the estimation, the proportion of the figures of the units, which delivered their results timely, to the total value during a period in the past was considered.

Several models of such type were analysed: T+15, T+20, T+25, T+30 and T+35 for the time lags, 6, 12, 18 and 23 months for the calculation of the estimation factors. Also post-stratification was considered, i.e. estimation of total values for different strata, which were defined by size.

For the analysis, data from 1997 to 1998 was used for an ex-post analysis.

The result of this analysis was quite disappointing since high error rates were observed using these models. Several reasons were found to explain this behaviour. First, as the sample was selected non randomly, it was considered as highly non representative. Estimations based on such samples are generally biased. Second, the calculation of the estimation factors could not be worked out consistently for all units within the sample because the required information was not available for all units. Again, biased results may be the consequence. This also, because shorter periods are more influenced by seasonal effects than larger periods. In contrast, shorter periods are more adaptive for the (short term) economic trend. Third, high fluctuation rates within the sample from one month to another was observed. This results in a general loss of significance for calculated change rates.

It was astonishing, that the dependence of the estimation quality on the time lags considered were not as expected. Even for T+35, which is very close to the final estimate at T+40, high error rates (revisions) were calculated. In contrast, the introduction of strata helped a lot to reduce the error terms. But still, the calculated revision terms were considered unacceptable for the application of such a method for German official statistic.

- b) For timely annual results, an estimation of the figures for the last months November and December are undertaken regularly.

These estimated yearly results are published at the end of the observed year, i.e. at T+0. The first yearly figures are available at T+42.

In fact, to obtain the estimated figures, three different estimation (nowcasting) methods are applied to deliver three results. These methods rely on moving average and more general ARIMA models, where the parameters of the models are recalculated every year. At the end, the plausibility of the figures are discussed among the experts within a meeting and it is decided in the meeting upon the figure to be published by the end of the year.

Concerning revisions, for 2002 the estimated value for the German exports was 647 Billion Euro and for the imports 520 Billion Euro. The corresponding final estimates were 648,3 and 522,1 Billion Euro, respectively.