



Economic and Social Council

Distr.: General
14 June 2023

English only

Economic Commission for Europe

Conference of European Statisticians

Seventy-first plenary session

Geneva, 22–23 June 2023

Item 9 of the provisional agenda

Timeliness, granularity and frequency of official statistics

Incorporating administrative data in monthly Labour Force Surveys estimation of economic activity, and Value Added Tax-based monthly estimation of foreign trade of services in Hungary

Prepared by Hungary

Summary

This document describes two new approaches adopted by the Hungarian Central Statistical Office to improve the quality and timeliness of the statistical data. The first part focuses on incorporating administrative data in the monthly Labour Force Surveys estimation of economic activity. The second part presents Value Added Tax-based monthly estimation of foreign trade of services in Hungary.

The document is presented to the Conference of European Statisticians' session on "Timeliness, frequency and granularity of official statistics" for discussion.



I. Incorporating administrative data in monthly Labour Force Surveys estimation of economic activity

A. Introduction

1. Due to rising user demands, notably following the rapid economic changes brought on by the COVID-19 pandemic and their effects on the labour force, the question of providing monthly Labour Force Surveys (LFS) data has become imperative for the Hungarian Central Statistical Office (HCSO). The Hungarian LFS is a continuous data collection, the quarterly sample of which consists of three statistically independent monthly sub-samples with large-scale fluctuation. Because of that, the reliability of monthly data is low, so we had to create a method which has to follow the economically explicable monthly changes while not being distorted by the fluctuation of the independent monthly samples. In order to satisfy these needs, with the help of administrative data the HCSO started to develop a monthly estimate for employment and unemployment numbers in Hungary.

B. Administrative data sources

2. While exploring the available administrative data sources, we had to keep in mind that there were certain criteria to be fulfilled: since the final data had to be available in certain sex and age subgroups, we needed administrative data available in these subgroups as well, for it to sufficiently guide the model estimation. The data used had to be available preferably in a long time-series for back-calculations, as observing a longer time period could help us ensure the lasting reliability of the estimation process.

3. Additionally, a challenging aspect that limited our choices was timeliness, as we needed monthly data that was also available on time for our first release (which usually falls near the end of the month following the reference month), so it can include the model estimation results. This narrowed down our options to the tax records of the National Tax and Customs Administration for the estimation of employed persons, and registered jobseeker records of the National Employment Service for the estimation of unemployed persons.

C. Employment

4. For the estimation of the number of employed persons we ended up using a combination of different tax records; employee records, self-employed records and so-called “small taxpayer” records. In terms of the aforementioned demographic dimensions, the data we had was available in suitable age and sex subgroups, however the processing of the records takes relatively long (almost two months following the reference month). To combat this, another administrative data source had to be included, which records the monthly inflow and outflow of citizens into the Hungarian social security system through employment. We used these records as supplementary data to predict tax data in advance. This prediction is later revised with the actual tax data, once it is available, but through preliminary research we have found that the prediction of supplementary data on tax data is quite accurate.

5. While there is obviously a sizeable overlap in the LFS and the tax records, we can presume that there are certain limits to the two data moving together as there are differences in the populations and concepts covered between the two data sources. LFS only covers residents of Hungarian private households, so people living in certain institutions or abroad – but still working in Hungary – can appear in the tax records, while they are not part of the observed population of LFS. On the other hand, certain groups could be missing from administrative data but be observed by LFS, such as residents of private Hungarian households that work abroad.

6. Aside from the conceptual differences of the observed groups, we also had to take into consideration the conceptual differences of the indicators used, as LFS uses average monthly estimates while administrative data gives us the exact number of people in the registry in the

reference month. This can become particularly troublesome while accounting for short-term, seasonal jobs.

D. Unemployment

7. For the unemployed, we decided to use data on registered jobseekers to guide the unemployment estimate of LFS. However, a considerable group of the people in the registry might not satisfy the ILO concept of unemployment. A prime example of this is people being able to do casual work, which the National Employment Agency allows people to do while being registered as jobseekers, but a couple hours of work can already constitute as employment by ILO terms. Additionally, one of the most prominent groups of this conceptual disconnect includes people merely being registered and thus showing up in the database, but not actively seeking work. There is also the case of people considered unemployed by LFS standards but simply not being registered as jobseekers, as it is not compulsory to do so in Hungary.

8. Conceptual differences between the indicators were present in unemployment data as well, as administrative data records the number of people in the registry at a given date. Although, in this case the date is the 20th of the reference month, not the end of the month, which could cause additional discrepancies in the movement of the two data. Fortunately, timeliness is not an issue in this case, as registered jobseeker data is available even before LFS data of the reference month is processed and the time-series available dates back to the early 2000s, so back-calculations were feasible as well.

E. Model estimation

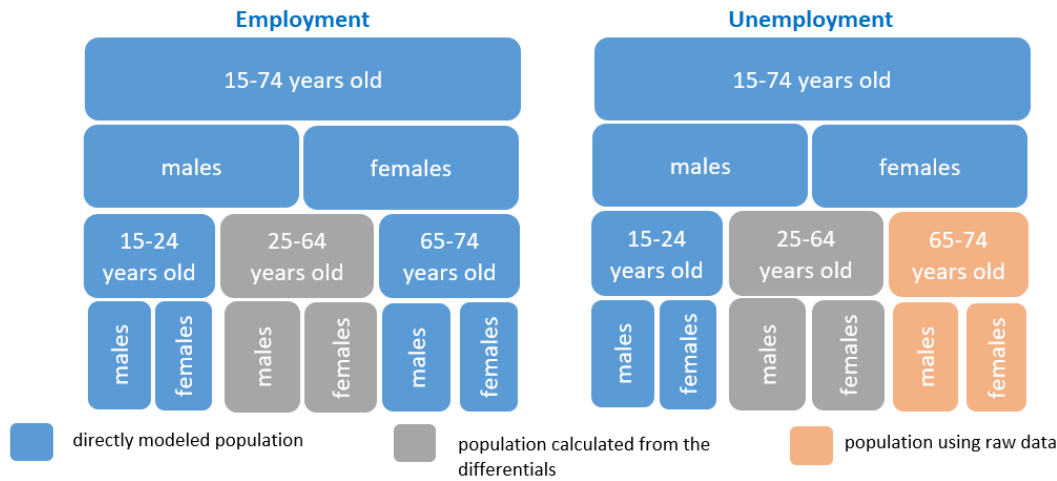
9. The Employment Statistics Section and the Department of Methodology worked together on the development of the new estimation method. After the preliminary research the Department of Methodology decided that the best method to incorporate administrative data would be through the so-called “state space” models, which can be used for regularly measured data such as the LFS. The basis of such models is that there are constantly changing, but directly not observable events, for which we would like to have an estimation. In our case, the directly not observable events could actually mean observable events, however these observations might not be the most accurate, such as our monthly employment and unemployment statistics derived from LFS.

10. To get a stable, functional estimation, state space models had two main supports: the tendencies and movements of the raw estimate’s error in all the needed demographical subgroups and the previously described administrative data sources. As the sampling method and the survey were well-known to us, most of our preliminary research and work was focused on the administrative data sources. Beyond the conceptual differences, we examined the time series of all administrative data and LFS data available, checking for any further discrepancies and other unusual shifts in the time-series, as well as whether these shifts had any underlying economic explanations or were something we had to handle as outliers.

11. During the estimation process, the 15-74 age group for both employed and unemployed persons was calculated first. The Methodology Department found that it is more effective to give an estimation for the total population observed first and to calculate the demographical subgroups afterwards, instead of determining subpopulations’ data first and their sum making up the total population estimate. Following the estimation of the 15-74 age group, 15-24 and 65-74 age groups were estimated and other needed subpopulations were calculated from the differentials of these three given age groups. However, for unemployed persons the 65-74 age group was too small to model, so we ended up using raw LFS data for the subgroup.

12. With the addition of the sex variable being included for each age group, we ended up with model estimates for 9 subpopulations for the employed and 6 subpopulations for the unemployed (see Figure 1). A core principle while developing the model was that the three-month moving averages of the monthly estimates should be close to the three-month moving averages of the raw monthly data, even in the aforementioned demographical subgroups.

Figure 1
Full and partial population estimates



F. Results

13. The three-month results of the model confirm that, it is possible to produce better quality estimates by including administrative data. With the help of state space models based on LFS data, tax records and registered jobseeker records, we ended up with results that were more consistent with other administrative data published by the HCSO and thus less confusing for the users. In addition, the model controlled the volatility of the monthly sub-samples better, while it was able to follow real-life changes, like the impact of the COVID-19 restrictions on the labour force market in the spring of 2020.

14. It is clear that, the estimation of both the employed and the unemployed not only effectively “smooths out” the fluctuation of the LFS subsamples, it also effectively follows seasonal monthly trends of the administrative data in a certain manner, that LFS three-month moving averages are not able to trace (see Figures 2 and 3). By combining the two data sources, it becomes possible to examine labour market processes and their underlying economic causes in a more complex way.

Figure 2.
Number of monthly employed people

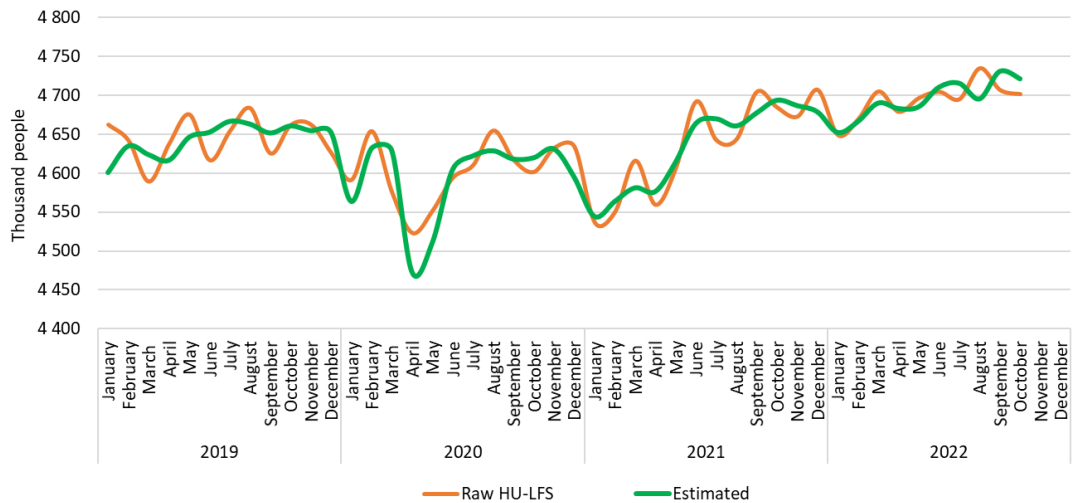
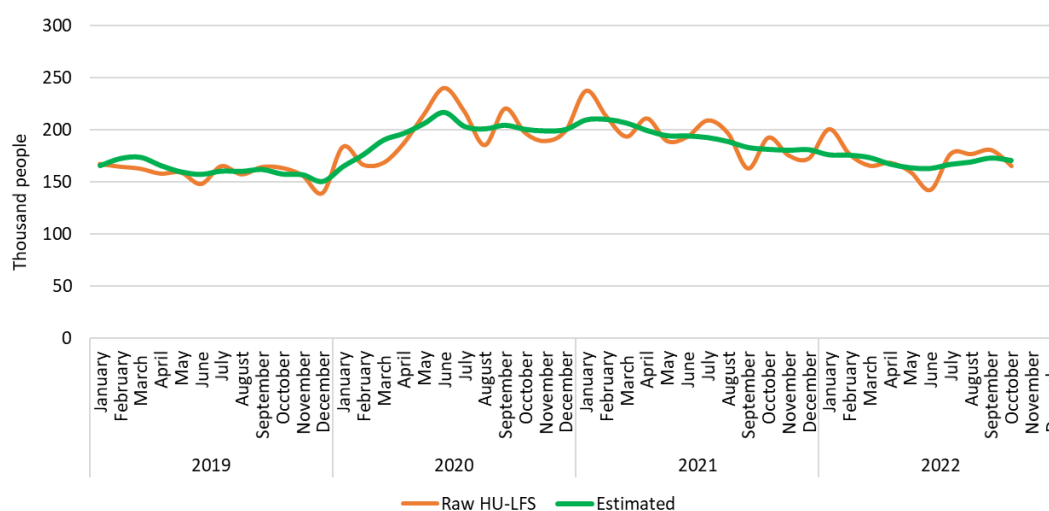


Figure 3
Number of monthly unemployed people



G. Publication

15. The methodological change was introduced as part of a major revision on February 24th with the monthly first release of January 2023. The published monthly time series were replaced with the results of the model on the HCSO website, and the revised time series were sent to Eurostat as well. As a result, we managed to comply with our legal obligation regarding the monthly unemployment rate, while providing data necessary to create and monitor the European Union labour market policies.

16. We back-calculated the data until 2011, which caused a break in the time-series, as for the years before 2011 we only have had the raw monthly LFS data available. After the initial publication, we decided not to revise the time-series monthly (mostly to avoid confusing the users), we only extend it by adding the latest results as the time-series develops. It is important to highlight, the three-monthly moving average and the quarterly data remained the main indicator in HCSO's Employment Statistics publications, the monthly data is provided as complementary information with a higher degree of precision than before.

H. Summary

17. For the past couple of years, the HCSO had been in the process of developing the model estimation for monthly employment and unemployment data. As a result of the estimation, the available monthly data became suitable for temporal comparison parallel to the improvement of the other quality criteria (relevance, accuracy, accessibility, comparability).

18. Information from several data sources can provide a more precise picture of the labour market. The estimation procedure demonstrated to us that the combined use of a population survey and administrative data is a promising method in official statistics.

II. Value Added Tax-based monthly estimation of foreign trade of services in Hungary

A. Introduction

19. In the Hungarian foreign trade statistics, the foreign trade turnover of services is measured on the basis of turnover data from quarterly data collection. In order to improve the timeliness of the statistical data, based on the monthly VAT data, we use a regression method to estimate the level of the monthly service-foreign trade turnover. With the help of the estimate given for the monthly frequency of foreign trade in services, a monthly flash indicator for foreign trade in services is produced, which significantly improves the timeliness of foreign trade statistics, which thus also provides great help in determining the monthly balance of payments data more precisely.

B. Historical background of trade of services statistics

20. The purpose of foreign trade product turnover statistics is to monitor product movements that physically cross the border as a result of transactions between resident and non-resident units. Any activity that represents a mutual obligation for services between a resident and a non-resident is classified as foreign trade in services. In the Hungarian statistical system, between 1992 and 2002, the Ministry of Economy and Transport collected data from designated data providers.

21. However, due to the coverage problem (in terms of value, type of service and geographic units), this resulted in low-quality data. Between 1996 and 2008, the official statistics were supplemented with settlement data from the Hungarian National Bank, according to the Balance of Payments Manual (BOP Manual 1993). In parallel, in 2002, an agreement was reached between the two institutions regarding the production of service foreign trade statistics. The Hungarian Central Statistical Office set up a new system with the support of the Hungarian National Bank. 38000 companies were selected based on the following data sources:

- Foreign trade data collection service of the Ministry of Economy and Transport;
- ITRS data from the MNB;
- VAT export and import data;
- Customs data.

22. After 2005, the Hungarian Central Statistical Office became responsible for the production of service foreign trade statistics. International trade players (directed sample) provided the majority of respondents. High-quality registers, appropriate sampling techniques, as well as the planning of questionnaires and data processing were also necessary for the production of statistics of adequate quality. After the establishment of the data collection and data production system, the Hungarian Central Statistical Office provides data related to foreign trade in services for both balance of payments statistics and national accounts.

C. Emergence of the need for a monthly estimate

23. In the current statistical system, the balance of payments statistics produced by the Hungarian National Bank publishes monthly data, which is prepared by continuing the quarterly data based on a model and then breaking it down into monthly frequencies. However, the statistics produced in this way have a high degree of inaccuracy, in the short term due to COVID, the forecast does not work and the estimate does not handle seasonality within a quarter.

24. Based on all of this, the professional demand was formulated that service foreign trade data should be broken down into roughly homogeneous components, which components can be estimated with monthly frequency data.

D. Model estimation and results

25. In order to improve the timeliness of the statistical data, based on the monthly VAT data, we use a regression method to estimate the level of the monthly service-foreign trade turnover.

26. Among the distinct main types of services, the greatest emphasis was placed on the VAT-based estimation of the service fees for contract work, transport services and business services. In the case of business services, the estimation is basically done with the help of monthly VAT data, based on the examination of the VAT data of the reporting companies and the correlation relationships. The partial estimate proved to be stable, with an explanatory power of around 80 per cent.

27. In the case of transport services, we treat air transport services separately, because in the case of air transport, administrative data related to airport traffic is also available for estimation. In the case of road transport, we were able to use the camera data to supplement the VAT data, in order to be able to give a more accurate estimate. Regarding the foreign trade turnover of air transport and related ancillary services, the correlation studies showed that there is a close relationship both with the number of transported passengers and with the product turnover. The partial estimate proved to be stable, with an explanatory power of around 90 per cent.

28. However, the available VAT data do not fully cover the information base of the quarterly data collection. Thus, an additional estimation is going to be added to the actual estimation, based on non-VAT data completes the estimation procedure approximating the total turnover. One of these additional elements is export and import turnover for tourism purposes, which is approximated with the data of the monthly tourism performance statistics, and on the basis of bank card data, with the help of the data of spending for tourism purposes. Supplementing the model with an estimate of tourist services can already make the estimate suitable for publication, as the foreign trade turnover of services is completely covered.

E. Summary

29. In order to improve the timeliness of the statistical data, we used a regression method to estimate the level of the monthly service-foreign trade turnover. With the help of the estimate given for the monthly frequency of foreign trade in services, a monthly flash indicator for foreign trade in services can be produced, which can significantly improve the timeliness of foreign trade statistics, which thus also can provide great help in determining the monthly balance of payments data more precisely.

Figure 4

Exports and imports of business services

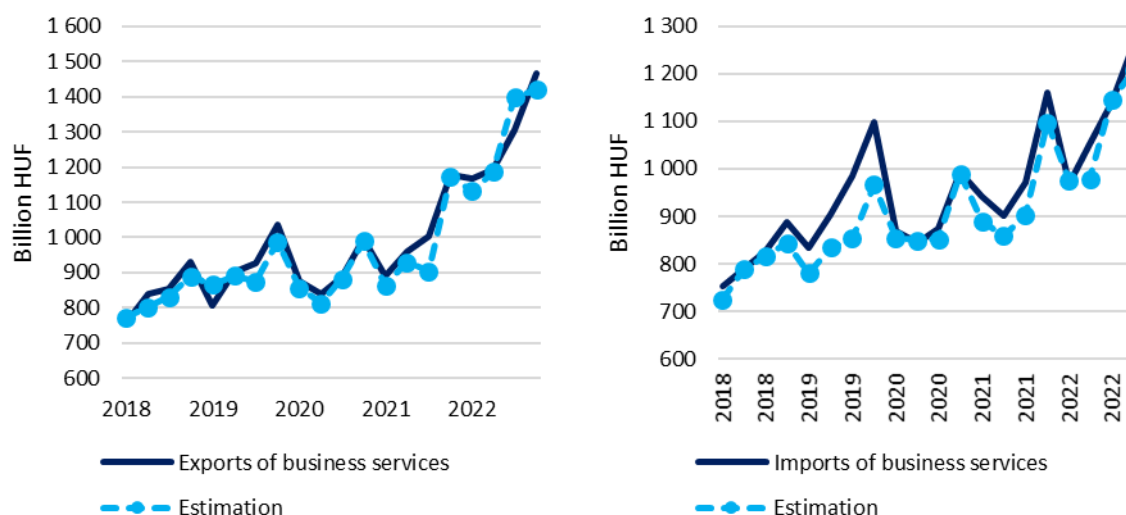


Figure 5
Export of services

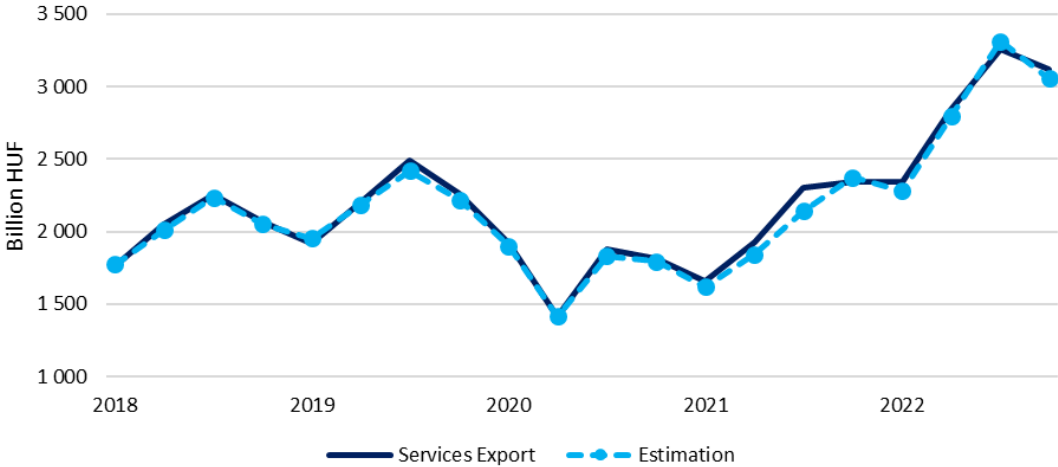


Figure 6
Import of services

