Integrated Software Tool for Processing Accountancy Data Information at EU level – An Application of GGIG

Sebastian Neuenfeldt, Alexander Gocht

Johann Heinrich von Thünen-Institut
Institut für Ländliche Räume
Bundesallee 50
38116 Braunschweig
sebastian.neuenfeldt@vti.bund.de
alexander.gocht@vti.bund.de

Abstract: The FADN (farm accountancy data network) provides farm accountancy data to determine and analyse income of agricultural holdings. The underlying data base has to be prepared for properly assessing the impacts of EU policy changes and market developments on agricultural income. This article presents the FADN data mining tool that reads and converts the given data for analysing the data itself and processing the data for further analysis. This tool is extendable and usable without software development skills, is free and open source and handles large data sets in an efficient and user-friendly way. It applies GAMS for aggregation, comparison, calculation, as well as for estimation of desired indicators that feed economic models.

1 Introduction

This article is based on the deliverable 4.1 [Ne12] of the EU project "FADNTOOL - Integrating Econometric and Mathematical Programming Models into an Amendable Policy and Market Analysis Tool using FADN Database". It aims to assess the economic impacts of EU policy changes and market developments on farmers and the whole economy. The FADN provides accountancy data to determine income and business analysis of agricultural holdings. The data base represents about 6,400,000 farms for the EU-27 and is a source of micro-economic data, which is harmonized using bookkeeping standards. Approximately, 1,000 variables are collected and transmitted by each member state (MS). The data comprises physical and structural (crop area, livestock numbers, etc.) data as well as economic and financial data (sales, assets, subsidies, etc.). The aim of this article is to present the developed software concept, which should first process and read the FADN data base, second, enables the user to convert and aggregate the data suitable for building up estimation approaches and mathematical programming models for analyzing supply behaviour in the EU. The article starts explaining the structure and coverage of the utilized FADN data set to understand the software concept requirements. We proceed with an overview of the FADN data mining tool and draw some conclusions.

2 FADN accountancy framework and underlying data base

The FADN was constituted in 1965 (Council Regulation 79/65). Derived from national surveys, FADN is the only source of micro-economic data that is harmonized using bookkeeping standards across the EU-27. The FADN sample does not cover all agricultural holdings, but only those, which, due to their size, are considered to be commercial. The methodology applied attempts to provide representative data along three dimensions: FADN region, economic size, and type of farming. The accounting and recording standards of the FADN are specified under EU regulations, but the data is collected by MS organizations.

The underlying data base consists of the EU-27 MS for 1990-2008. Since 1990, 274,000 farms contributed to the network in different years. Table 1 summarizes some indicators obtained from the FADN data base. In 1990, 57,615 farms represented a population of 4.15 million farms. These numbers increased due to the enlargement of the EU to 78,137 farms and a population of 4.95 million farms. The ongoing process of structural change is reflected by the declining number of farms if corrected by the enlargement effect. As a result, the average farm size increased from 20 ESU1 in 1990 to 30 ESU in 2008. This figure fluctuates at the aggregated EU level, due to the inclusion of Austria, Finland, Sweden and East of Germany in 1995, the enlargement in 2004 by the EU-10, and in 2007 by the EU-2. It is further influenced by the continuous increase of applied thresholds, defining the minimum size of a farm considered in the sample, which varies according to the agricultural structure from one country to another.

	1990	1993	1996	1999	2002	2005	2008
Farms represented	4,152,997	3,834,367	3,844,146	3,669,180	3,042,444	4,136,547	4,954,812
Sample farms	57,615	56,529	58,347	58,599	58,487	76,555	78,137
Average ESU	20	23	27	31	38	33	30
Utilised agricultural area (UAA) in million ha	93	93	111	111	107	139	158
Rented UAA % share on UAA	47	46	50	50	52	51	53

Table 1: Summary FADN data base for the EU-27 over selected years

3 The FADN data mining tool

The data of the FADN data base is provided in a comma separated values (CSV) file format by DG-AGRI. Each file comprises data for a certain country and year. The FADN data mining tool aims to test and verify the quality and consistency of the underlying data base. This requires an integrated software tool, which can be extended and used by project partners without software development skills and should therefore be

---

[1] European size unit, abbreviated as ESU, is a standard gross margin of EUR 1 200 that is used to express the economic size of an agricultural holding or farm. (see
http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Glossary:Economic_size)

free or open source and handle large data for processing and comparison as well as to be capable to use GAMS for extraction, aggregation, calculating of indicators and estimating parameters for mathematical programming models.

These requirements are met with the following software tools. A publicly available Java program transforms each CSV file into a GDX file format [Go09] which is the input (and output) file format for the FADN data mining tool. The structure of the FADN data mining tool is illustrated in Figure 1. It is based on the GAMS Graphical Interface Generator (GGIG) [Br10, Br11], that generates a graphical user interface (GUI). The user operates the GUI which uses text files that have a project specific layout. The GGIG and the GUI start user-defined GAMS programs and generate GAMS compatible meta data.
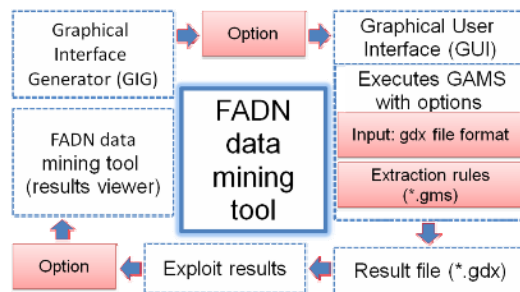

Figure 1: Illustration of the software tools integration

The resulting files are written in the GDX file format and can be viewed in a user-friendly and efficient way in the FADN data mining tool (results viewer) in predefined tables (Figure 2), figures and maps of NUTS II regions for MS or the EU. This tool is also capable to compare the extracted data between regions, years or generated indicators.

Figure 2: Screenshot of the predefined tables (Source: FADN data mining tool)

The user can access the original FADN data base for all countries and years for a selection of variables as a "FADN exploitation task". The user defines which information is needed and starts the task "extraction rules" with the program (GAMS file), in which the extraction rules are implemented. The extraction rules are closely

defined to the information needed for the mathematical programming models and "transform" the data from the accountancy tables of FADN data base into the parameterization of farm models. Therefore, indicators for crop and animal production activities, costs, subsidies and income are extracted. The GDX results file is displayable and capable in the results viewer. To provide an overview of the rules and to detect possible problems in the data, the FADN data mining tool also aggregates, in addition to the single farm level, all indicators at different regional levels such as farm type, ESU, NUTS II, MS and EU. For an analysis at the farm level, a heat map was implemented in the FADN data mining tool together with a ranking routine, which can be used to analyze the complete sample of farms of a certain indicator over time.

Although we can apply all the extraction rules at farm level, the resulting 1.7 Giga Byte file cannot be loaded and viewed in the results viewer. To avoid this, a separate file is written with all information for all countries and years excluding farm level information. This can be used to analyze the effects EU-wide. Detected problems can then be analyzed either for a certain year or MS. To process the extraction rules in an acceptable execution time, parallel processing was applied and a run for all farms, countries, and years takes less than 1.5 hours. All the results are stored in a GDX file format, which can easily be accessed as input by other partners. The predefined views in the FADN data mining tool are structured to allow the data to be analyzed by pivoting, sorting, and applying descriptive statistics.

4 Further developments and conclusion

Within the FADNTOOL project it is foreseen to extend the FADN data mining tool with a work step component, which estimates variable input costs. This component takes the results files as input data. Furthermore, it is planned to host the farm simulation engine within the setting. Because the gathered information of each user is needed to solve economic issues, the data mining tool is intended to be also capable to use output from other software (econometric packages, etc.) and to produce input for other software.

In summary, the FADN data mining tool is intended to guide and support the user in selecting the desired data to be analyzed, in proofing the consistency, in viewing the extracted data in lists, figures or maps and finally in using the data for further analysis like mathematical models.

References

[Br10]   Britz, W.: GGIG Graphical Interface Generator. Technical documentation. Bonn, 2010.
[Br11]   Britz, W.: The Graphical User Interface for CAPRI – version 2011. Bonn, 2011.
[Go09]   Gocht, A.: FADN to GDX. Technical documentation. Braunschweig, 2009.
[Ne12]   Neuenfeldt, S., Gocht, A., Karagiannis, G., Stefanou, S.: A Handbook on the use of FADN Database in Programming Models. FADNTOOL Project. Deliverable 4.1. Braunschweig, 2012. (forthcoming)