

Konadu, Magneta; Kalthoff, Herbert

Article

The infrastructure of financial markets: The case of statistical information

economic sociology_the european electronic newsletter

Provided in Cooperation with:

Max Planck Institute for the Study of Societies (MPIfG), Cologne

Suggested Citation: Konadu, Magneta; Kalthoff, Herbert (2013) : The infrastructure of financial markets: The case of statistical information, economic sociology_the european electronic newsletter, ISSN 1871-3351, Max Planck Institute for the Study of Societies (MPIfG), Cologne, Vol. 14, Iss. 3, pp. 34-40

This Version is available at:

<https://hdl.handle.net/10419/156017>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Infrastructure of Financial Markets. The Case of Statistical Information

By **Magneta Konadu and Herbert Kalthoff**

University of Mainz

konadu@uni-mainz.de

Herbert.kalthoff@uni-mainz.de

Introduction

Recent sociology of economic knowledge practices – Social Studies of Finance – has intensely examined the situational and technological fabrication of international finance sectors in the last years. In the course of this, diverse phenomena came into view, such as trading on FX-markets, the mechanism of price formation in the stock exchange, the development and implementation of formulas and also the organizational and interactive setting of financial trade (for example Knorr Cetina/Brügger 2002; Beunza/Stark 2004; MacKenzie/Millo 2003; Preda 2009; Muniesa 2000). Conceptual recommendations formulated within this research have (a) stressed, from the perspective of the science and technology studies, how economic (theory) models enable economic practice, and thereby ascribed a performative power to these models, (b) described, from a micro-sociological point of view, the technological apresentation supporting and actuating the trade apparatus and (c) drawn the attention to the performativity of those calculative and representing devices actors use in order to implement transactions.

In the course of this discussion the role of (technological) infrastructure supposedly constituting trade was often emphasized. It is this infrastructure that makes trade in its rapidity (time dimension), in its intricacy (material dimension), and in its cooperative accomplishment (social dimension) possible. Repeatedly, it was also investigated what exactly is represented mathematically or economically through technology (see for example Kalthoff 2011; Kalthoff/Maeße 2012; Lépinay 2011). In these studies, technology is considered as a medium facilitating other media. Representation on the surface of the screen happens through alphabetical or mathematical script, signs or symbols. This means that technical artifacts blank out their own texture in order to make the other visible. The re-

search at hand responds to the relevance of the depicted: Watching finance traders as subjects of financial trade, one sees that it is exactly this visualization of economic information which matters for them. The notion of information is somewhat colorful in these terms: “information” is viewed as an entity bearing an omnipotent power. One needs to possess it (manifold) in order to make reasonable and foresighted decisions. This is linked to a second aspect: “information” seems to embody vividness and transparency; owning it enables decision-making. Both aspects suggest a transparent intentionality, as everyone understands what is called “information” without further ado. The notion of information thus achieves a double status: on the one hand it becomes central to every sociological analysis of economic processes, because market actors constantly depend on obtaining and evaluating information; on the other hand the term is not sociologically examined, as simply postulating information is not sufficient.

This problem is twofold: firstly, information is not distributed evenly, but asymmetrically (Akerlof 1970) and it is often incomplete; secondly, there is no inherent evidence within information on how to interpret it, but it rather offers spaces for interpretation which result in different possible decisions. Third, and last, information is not accessible in the pure (“naked”) form, but is always bound to forms of expression such as tables, diagrams, scripts or charts. This is to say that information always appears in a medium formatting that what it depicts in its own way. „Acting on the market” or „acting economically” thus means having displays storing information at one's disposal and being able to decode these displays.

Following the studies outlined above we will look at one case of the creation of statistical information, which – besides much other information – is (or can be) used by finance traders. We address the issue of statistically determining household finance in Europe; the material we show was generated during 18 months of ethnographic fieldwork at the Deutsche Bundesbank (German Central Bank). We will not, however, go into detail about the ways in which these and other financial statistics are used by finance institutions.

1 Financial Statistics as Survey Statistics

Surveys are, beside experiments, observations, polls, and administrative record keeping, a modern tool for gathering valuable information about large populations. Surveys are mostly implemented to collect facts about social and economic matters. There are ample surveys on various topics that are conducted on a regularly basis. There are surveys on unemployment, crime, assessments of educational performance and consumption, which give answers to questions that are highly relevant for society as such. In surveys, data is collected, processed and analyzed by research groups, which often consist of survey methodologists. Surveys are not conducted arbitrarily but rather follow a systematic set of applicable methods and theories that frame the inquiries and interviews that are performed. A major characteristic of surveys is the fact that they produce statistical, thus quantitative descriptions of populations. The construction of figures by appliance of statistical methods has to be understood as a procedure of making numbers valid. In order for numeric descriptions to be exact and the data to have a high quality, sources of error must be identified. In the context of surveys errors can be understood as "deviations from the true values applicable to the population" (Groves/Floyd/Coupler 2009: 3). The existence and longitude of panels prove that there are people, groups and institutions who are convinced that surveys provide relevant information and are therefore important. Survey statistics are used as data basis to "prepare reports, policy recommendations, scholarly publications, testimony for congress and documentation for use in court." (Groves/Floyd/Coupler 2009: 14)

Recently, the interest for surveys on household finance has increased in the economic field. Besides corporations and financial market agents, private households and their finances have become the center of attention. Economic researchers have recognized that private households face new challenges and possibilities due to several factors such as massive effects of financial crises, demographic changes (aging society) and retirement provisions, new financial products being installed and innovative investments emerging. How do different households react to sudden incidents (e.g. financial crisis) and institutional changes? What kind of assets do private households invest in? How diversified are their portfolios? How large is the financial burden which has been placed on households by transmitted loans? These are only some questions among many that can be answered by household surveys on finance.

A scientific study of private household finances that collects information about the above mentioned entities and novelties has been introduced by the European Central Bank (ECB). The primary goal of this European survey is to create a comprehensive picture of the financial situation in private households and to be able to compare results between Euro-countries. Included is information on wealth (e.g. money, estate), assets, income, insurances, inheritance and employment. The study generates a detailed dataset for scientific research and a basis for monetary policies made by central banks. A survey design drafted by researchers of the project "Household Finance and Consumption Survey" (HFCS) was adapted by the participating countries of the Eurozone. "Core-questions" and methodological steps are the same in all countries in order to assure comparability among the participating countries. With this endeavor micro-data about households can be used to complete already existing aggregated data. Moreover a thorough analysis of the financial portfolios of private households can be undertaken due to the detailed structure of the questionnaire. Household surveys in general consist of the sum of decisions that need to be made during survey design, data collection, data processing and data analysis. These multiple and interconnected decisions are made with the intention to preserve the quality of a dataset by identifying and minimizing anticipated mistakes and errors:

"The total survey error approach means taking that broad perspective and ensuring that no feature of the survey is so poorly designed and executed that it undermines the ability of the survey to accomplish its goals. [...] When a well-trained methodologist makes these decisions, it is with a total survey error perspective, considering all the implications of the decisions that are at stake and how they will affect the final results." (Groves/Floyd/Coupler 2009: 34)

The construction of the financial household survey of ECB is based upon the survey error approach since all activities and procedures are reflected in terms of foreseeable mistakes that should be avoided. The researchers have a strong and pronounced "consciousness of error" which leads them to implement search engines in order to detect different kinds of errors and inconsistencies. Finding mistakes, recognizing patterns and identifying interconnections of variables require an "immersion" into the micro-data. The micro-dataset is the starting point for a proper understanding of the unique character of the entire data. A bottom up perspective is applied in order to get a grasp of the complex dataset.

2 The Working Steps of Survey Construction

The fabrication of a household survey on finance is a process which implies many work steps. Each step has its own exigency and questions which need to be met. The working stages indicate the necessary “work” which is required in order to establish financial statistics based on household surveys. Like other forms of quantification¹ survey statistics involve effort, coordination and different kind of resources such as time and money. “Quantification requires considerable work, even when it seems straight forward.” (Espeland/Stevens 2008: 410)

See appendix, Figure 1

The process that has to be followed begins with the survey design which partially consists of drafting a questionnaire, pre-testing and evaluation of questions, creating an outline of decisions, which have to be made in the course of construction and finding solutions to the following questions: What is the target population? How are the respondents selected? (sample design), How is the data collected? (mode of data collection). Thereafter follows the selection of the sample. Household surveys currently use “area probability sampling” to select households. This means geographic areas are sampled as clusters by listing the housing units in these areas and providing all households with the same chance to be sampled. The collection of systematic data is the next step. Historically, gathering information was first done by means of paper questionnaires in state census since they were considered a low cost option for measuring populations. Today the means of data collection have broadened with proliferate use of telephones and the invention of computers. Beside pen and paper interviews (PAPI), there are, amongst others, computer assisted personal interviews (CAPI), computer assisted telephone interviews (CATI), and more recently computer assisted web interviews (CAWI). These methods of data collection vary particularly regarding the degree of privacy, in-depth interaction with respondents, and interviewer involvement.

After collecting the data there are further set-ups that need to be made, which leads to a continuous processing of the dataset. The following actions appear after data collection (also Groves/Floyd/Coupler 2009: 330-331):

a) *data entry* – the entering of numeric data which has previously been transformed from textual answers into data files

b) *data editing* – the dataset is screened for errors and inconsistency that are discovered and corrected

c) *imputation* – missing item values are replaced by estimated values; it is a method of reconstructing missing data and a scientific way of handling item non-response

d) *weighting* – adjustments of computations that are made to compensate for unequal probabilities of sample selection, unit non-response and deviations of key variables of the population that are known from other surveys and censuses

e) *anonymization* or de-personalization – the data is transferred into a form in which based on given information the identification of survey respondents is impossible.

The focus of this paper is on the process of data editing since this is one of the most complex and at the same time delicate stages of data manipulation that is relevant for the accuracy of a survey data. Data editing secures “trust in numbers” (Porter 1995) by scrutinizing the entire dataset and by a rule based transformation of values. The entire process of survey construction might be linear in its design, but that does not exclude overlappings, circularities, and repetitions of segments. In complex household surveys data collection, data editing and preparations for imputation might all occur simultaneously since there are different groups (e.g. researcher, interviewer, coder), each assigned with a specific task to accomplish.

3 Data editing

The main goal of data editing is to increase the quality of the dataset by detecting and correcting errors and logical inconsistencies in the data. Data editing directly prepares for imputation and weighting. In essence it deals with the manipulation of recorded data to improve the condition of the data. With the use of software such as, for example, Stata the complete dataset is screened for errors and the dataset is restored. The correction of survey data generally consists of the following activities as observed in ECB’s survey on household finance: Assessments of information on conduct of interview, evaluation of interviewer comments on interviewee’s specifications, coding of open

answers, formal data checks, plausibility tests and outlier analysis.

Assessments of information on conduct of the interview

It is common in household surveys that paradata are generated at the end of the interview. The interviewer is requested to answer questions for example on the outward appearance and interior conditions of the home, and to comment on the interview process. Was the interviewee suspicious before and after the interview? How was the interviewee's understanding of questions? How do you judge the reliability of the information provided by the interviewee? Did the interviewee consult any documents to provide answers? How was the interest in the interview? These are some of the questions that are posed. The interviewer comments are analyzed as background information on the interview itself in order to get a general impression on the overall quality of answers. Another factor is the duration of the interview, which is also an indicator for the quality and thoroughness of the interview. All these pieces of information are analyzed and summed up in a short report including the editor's notes on complications. Afterwards the households are categorized as problematic or unproblematic cases.

Coding of open answers

"Coding is both an act of translation and an act of summarization." (Groves/Floyd/Coupler 2009: 332) It is a translation in the sense that it translates textual information into numeric data. Variables which differ qualitatively such as "highest vocational training" (categorical variable) and "income" (continuous variable) are commensurable – they can be related and compared to each other – because the same metric is used. Commensuration transforms differences into quantity. It unites objects and entities by encompassing them under a shared cognitive system (Espeland/Stevens 2008: 410). Coding also summarizes by assigning individual textual answers into one code category. Since the majority of the questions asked in a household survey are provided with code categories, there are only a few questions or variables which allow open answers that need to be coded a posteriori. In those cases the non-numeric answers have to be categorized and then attached to distinct numbers. There are other cases where the coding structure includes the option "others" which should be used exclusively when the answer does not match any of the given codes. However, experience has

shown that many of these open answers under "others" can be assigned to the actual codes.

Variable HF1240 Type of fund
What type of fund does your household have?

- 1- Funds predominantly investing in equity
- 2- Funds predominantly investing in bonds
- 3- Funds predominantly investing in money market instruments
- 4- Funds predominantly investing in real estate
- 5- Hedge Funds
- 6- Other fund types (specify)

For example, a respondent answers the above question with the statement that the household owns a "real property". The interviewer overlooks for instance code *4-Funds predominantly investing in real estate* or does not find the same wording in the given code structure, he will code the answer "real property" in *6- Other fund types*, although it definitely should be assigned to code 4. Wrong classifications can also occur for example with NACE and ISCO codes for the description of occupations since these are complex and differentiated.² The overall aim of checking pre-codes and coding open answers is to assign as many open entries as possible to a numeric code.

Evaluation of interviewer comments on interviewee's specifications

Interviewers have the possibility to comment on every question that respondents are supposed to answer. They can highlight on issues and aspects that are judged as relevant for noting. The annotations can be included into the CAPI in a text field that is hidden behind a button that can be clicked on by the interviewer. Important references on inconsistencies, wrong specifications, and elaborations on the given answers can be found in the commentaries. These interviewer comments can help in deciding whether it is necessary to edit a value or not. They can give clues on how to evaluate and comprehend a given answer by the respondents. Besides, they can be taken into account while doing the formal consistency checks. The question "how did your household acquire the main residence you own? Did you purchase it, did you construct it yourself, did you inherit it, or did you receive it as a gift?" can be answered by a respondent by mentioning "inheritance". The respondent might go on with an elaborate account of how he received the house some years ago as an inheritance from his grandparents. The interviewer will save this additional information as a comment, since it is an important example of intergenerational transmission of wealth. The

interviewer's comments on each question have to be re-processed by categorizing the entries and by computing a frequency distribution for every interviewer.

Formal data checks

Data editing is accomplished through different kinds of checks. Core activities in any household survey are the formal data checks. They comprise of a number of checks that are implemented in order to look for inconsistencies in responses. Stata codes are used to screen for errors that then are corrected, also in Stata, by recoding the actual answers.

Formal data editing has to be well organized because if it is not done with care and if it is not rule abiding it can lead to endless changes in the data, which reduces the quality of the dataset.

See appendix, figure 2

The code in figure 2 is an example for a filter-value check of the continuous variable "number of private businesses a household owns" (hq0210). The first stanza checks the values and range of the variable hq0210: it should neither be less than zero nor higher than 20. Besides the numeric digit other acceptable values are "don't know" (-1), "no answer" (-2) and "question filtered" (-3). If the rule is violated and other unacceptable values are falsely recorded Stata computes errors and lists all the IDs of the cases where an error occurred. The two stanzas bellow are both filter checks that look out for errors in the logical tree that brings forth the assessed variable (hq0210). The second stanza computes an error under the condition that a household owns a business partially or fully (hq0100==1) and someone in the household is either self-employed or has an active role in running the business (hq0200==1), but the variable on the number of businesses is filtered or does not have any value. This is a logical error in the sense that if hq0100 and hq0200 are answered with yes (1) then the variable with the question "how many private businesses does someone in your household own entirely or in part?" has to follow up. The last stanza is the opposite filter-check of the second stanza: if hq0100 and hq0200 are not answered with a "yes" (hq0100!=1/ hq0200!=1), meaning if a household does not own a business and does not specify whether there is someone from the household in charge of it, the question on the number of private businesses should be omitted and not be asked in these cases. These filter

errors and value errors can be ascribed to interviewer mistakes or CAPI-errors.

Plausibility tests

Plausibility tests are conducted in order to detect frictions regarding the content of the data. Quantitative single entries are compared among each other and with aggregated context information. The underlying question that has to be answered is how probable the occurrence of a particular event is. For example, assume that a check on "employment" and "income" leads to the finding that an unemployed person has a monthly income of 5000 EUR. This amount of money is quite unlikely for a non-worker. It appears as if a wrong amount has been reported, which might turn out to be an interviewer error. But a further look at the aggregated monthly income reveals that the unemployed receives revenues from apartments he or she rents out. In this case there is no need for recoding and changing the income value. Besides internal plausibility tests there are external tests that compare aggregates from the dataset with average values from external data resources. For instance information from the dataset on the size and price of dwellings can be compared to figures from other established household surveys. Major differences may cause an adjustment of the data. Plausibility tests lay the ground for weighting.

Outlier Analysis

Statistical outlier analysis is important for the identification of extreme observations that deviate strongly from the normal distribution of sample members. Methods often used to find outliers assume that variables that are assessed have a normal distribution. They identify observations considered to be implausible based on mean and standard deviation. With the help of outlier analysis rare combinations of attributes can be found. This is, moreover, relevant for the anonymization of the data because rare constellations of values can disclose a household and therefore violate the protection of data privacy.

Data editing facilitates the recognition of patterns and since it is an in-depth micro-analysis it also trains the "eye" for discovering errors and inconsistencies. The constant and repetitive pattern of data editing consist of a) discovery of an error via Stata code, b) rule based manual recoding of incorrect values, c) documentation of errors and elimination strategies and d) reiteration of checks to avoid new errors while changing the data. The Stata codes that

are applied to detect errors together with the statistic software can be perceived in a more abstract way as a microscope, with which entities are seen in a close up and different layers of an object become visible.

Altogether, editing work in household surveys is laborious and standardized by guidelines that have to be followed. It is sequential and based on rules. Some of the activities involved are repetitive and therefore easy to automatize. Nevertheless there are various stages in the editing process where techniques of objectification are not possible, but rather a keen sense for possible errors might be the effective way.

Conclusion

In the text above we have depicted the generation of statistical information, information that is used by financial institutions for their business. The goal was to open the black box of this form of economic research in order to find out how it is implemented, how errors and inconsistencies are detected and eliminated, and which methodological knowledge flows into them. Thus we were able to show how defective datasets are transformed into (allegedly) "valid" statistics. Hence, the "numeric" figures generated in the process of fabrication are social as well as statistical constructs in many ways. Using the example of financial data, we have examined how the knowledge financial actors draw on is generated. This means that we did not understand statistical displays merely as a resource for actors, but made these displays themselves subject of research. Making large statistical information available does not solve the problem of uncertainty, but it conveys which practices enter these statistics before further use by market actors.

Magneta Konadu, MA, is currently a PhD-Candidate at the University of Mainz where she is doing her doctorate in *Sociology of Financial Statistics*. Her research interest lies in *Economic and Financial Sociology*.

Herbert Kalthoff is Professor of Sociology at the University of Mainz. He published *Wohlerzogenheit. Eine Ethnographie deutscher Internatsschulen* (1997), *Ethnografie. Zur Praxis der Feldforschung* (2013 together with G. Breidenstein, S. Hirschauer and B. Nieswand), he edited *Fact and Figures. Economic Representations and Practices* (2000 together with R. Rottenburg and H.-J. Wagener) and *Sozi-*

ologie der Finanzmärkte (2012 together with U. Vormbusch).

Endnotes

1Quantification is understood in accordance with Espeland and Stevens as "production and communication of numbers" (Espeland/Stevens 2008: 407).

2NACE stands for systematic statistic categories of economic branches within the European Union. ISCO is the International Standard Classification of Occupation.

References

- Akerlof, George A., 1970: The market for "lemons": Quality uncertainty and the market mechanism. In: *Quarterly Journal of Economics* 84(3), 488-500.
- Beunza, David/David Stark, 2004: Tools and the trade: The socio-technology arbitrage in a Wall Street trading room. In: *Industrial and Corporate Change* 13, 369-401.
- Espeland, W. Nelson/Mitchell L. Stevens, 1998: Commensuration as a social process. In: *Annual Review of Sociology* 24, 313-343.
- Groves, Robert M./Floyd J. Fowler/Mick P. Coupler, 2009: *Survey methodology*. New Jersey: Wiley.
- Kalthoff, Herbert/Jens Maesse, 2012: Die Hervorbringung des Kalküls. Zur Praxis der Finanzmathematik. In: Herbert Kalthoff/Uwe Vormbusch (eds.), *Soziologie der Finanzmärkte*. Bielefeld: Transcript, 201-233.
- Kalthoff, Herbert, 2011: Un/doing calculation. On knowledge practices of risk management. In: *Distinktion* 12(1), 3-21.
- Knorr Cetina, Karin/Urs Bruegger, 2002: Global microstructures: The virtual societies of financial markets. In: *American Journal of Sociology* 107, 905-950.
- Lépinay, Vincent A., 2011: *Codes of finance: Engineering derivatives in a global bank*. Princeton: Princeton University Press.
- Muniesa, Fabian, 2000: Un robot walrasien. Cotation électronique et justesse de la découverte des prix. In: *Politix* 13(52), 121-154.
- MacKenzie, Donald/Yuval Milla, 2003: Construction a market, performing theory: The historical sociology of financial derivatives exchange. In: *American Journal of Sociology* 109(1), 107-145.
- Porter, Theodore M., 1995. *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Preda, Alex, 2009: *Framing finance. The boundaries of markets and modern capitalism*. Chicago: University of Chicago Press.

Appendix

Figure 1

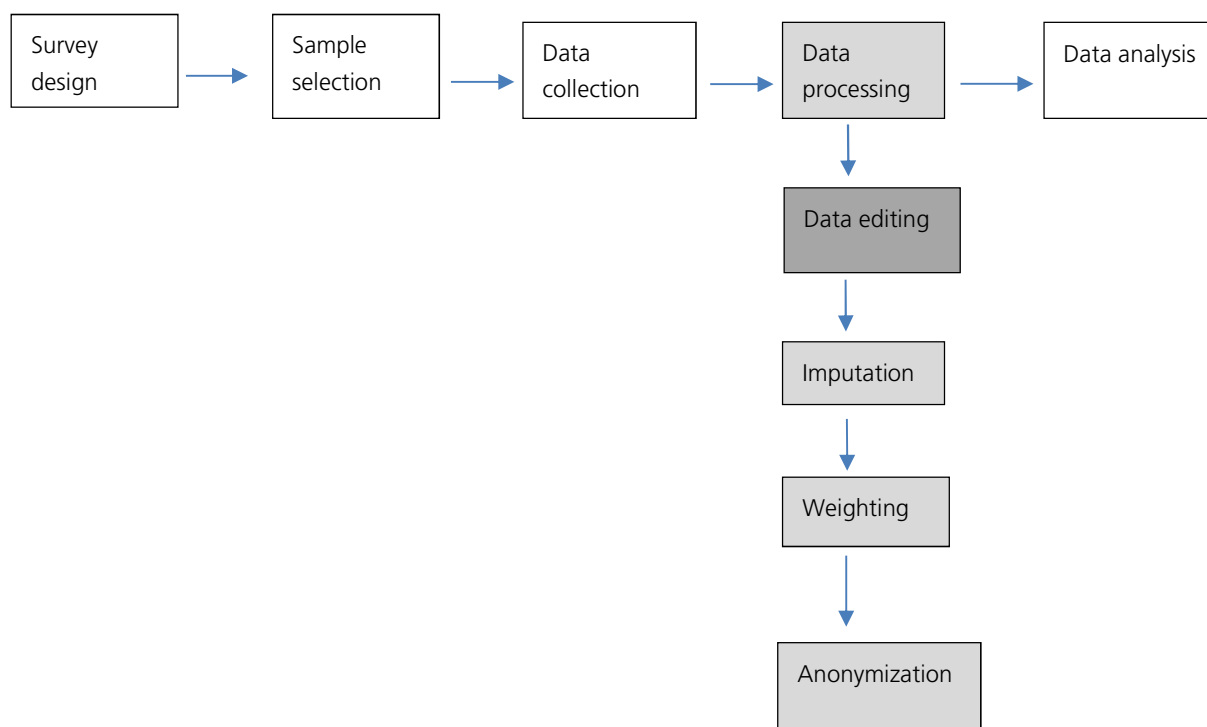


Figure 2

Filter-value check of variable hq0210 (number of private businesses)

* -----

```
replace fehler=1 if ((hq0210<0 & hq0210!=1 & hq0210!=2 & hq0210!=3) | (hq0210>20 & hq0210!=.))
```

```
tab fehler
```

```
list caseid hq0210 if fehler==1
```

```
replace fehler=0
```

```
replace fehler=1 if ((hq0100==1) & (hq0200==1)) & (hq0210==. | hq0210==3)
```

```
tab fehler
```

```
list caseid hq0210 if fehler==1
```

```
replace fehler=0
```

```
replace fehler=1 if ((hq0100!=1) | (hq0200!=1)) & (hq0210!=. & hq0210!=3)
```

```
tab fehler
```

```
list caseid hq0210 if fehler==1
```

```
replace fehler=0
```