DEUTSCHE BUNDESBANK EUROSYSTEM

Panel on household finances

2. PHF Scientific Data User Workshop

Martin Eisele (Statistics Department)
Junyi Zhu (Research Center)

April 16th 2013

Overview

Panel on household finances

Part I:

- **Basic facts about the PHF**
- **Workflows and methodology (what happened with the data?)**
- **Data structure and data content, variables**
- **Data application procedure**

Part II:

- **Using the PHF with Stata in consideration of:**
  - **Data preparation**
  - **Multiple imputation**
  - **Sampling design**
  - **Variance estimation**

PHF – a new Survey on Household Finances

Panel on household finances

The German PHF…

… is an integral part of the HFCS, the system of Euro Area surveys on household finances

… is a fascinating scientific endeavor in its own right, covering Germany

HFCN Surveys

Panel on household finances

2006: Establishing of a network (**HFCN**) at the ECB involving representatives of all Euro Area central banks

2008: Governing council of ECB decided that the HFCS will be conducted in all Euro Area countries.

- **HFCN**: harmonized **Core, mandatory and common** to all surveys
  - Main focus on „Household balance sheet“ (incl. debts)
  - Consumption patterns (basic information)
  - Socio-demographics
- **HFCN**: harmonized „**periphery“**, voluntary
  - e.g. questions on the impact of the financial crisis, financial literacy
- **Country-specific** components
  - Savings behaviour for pensions and other contracts

Household Balance Sheet

Panel on household finances

Assets	Liabilities
Non-financial assets – Owner-occupied housing – Other ownership of homes and property – Established businesses (net value) – Vehicles, collections, jewellery etc	*Liabilities* – Mortgages – Consumer loans (including credit card debt, current account credit, unpaid invoices, student loan debt) – Loans for business activity
Financial assets – Savings and current accounts, savings under building loan contracts – Mutual fund shares/units, debt securities, shares, derivatives and certificates – Balances from private pension and life insurance policies – Long-term equity investment – Assets under management	*Net wealth*
Total assets	Total liabilities

What is special about PHF

Panel on household finances

PHF gives special emphasis on key topics in German economic policy:

- **Savings**
- **Pensions and old age provision**
- Some special topics (e.g. „Bausparverträge“, self-assessment)

PHF takes a **life cycle perspective**, establishing a **full panel.**

Approach to savings borrowed from **SAVE**, self refreshening panel structure borrowed from **SOEP**.

Sampling

Panel on household finances

Basic facts:

- Population register based sampling
- Stratified random sample
- Oversampling of wealthy households

Target population: All private households located in Germany except institutional households (in old-age homes, prisons etc.)

Three first stage Strata: Wealthy small municipalities (30%), other small municipalities (30%), large cities (more than 100000 inhabitants) (40%).

Implicit geographical stratification:

In every stratum, municipalities are drawn with probability proportional to size with implicit stratification by region and municipality size class.

Oversampling of wealthy households

Panel on household finances

Strategy for small municipalities:

Use of income tax statistics to identify municipalities with high share of people paying high income tax (8% in total population, 30% in sample)

Strategy for big cities with more than 100000 inhabitants:

Use of micro geographical information (quality of residential area, type of dwelling, purchasing power indicator) to identify wealthy street sections (14% in total big cities population, 52% in the big cities sample) -> second stage stratum.

-> It worked!

Distributions of households by weighted net wealth deciles

1	2	3	4	5	6	7	8	9	10
7%	7%	7%	8%	7%	8%	8%	11%	15%	22%

Sample size

Panel on household finances

- **Second stage sample:** 230 **sample points** (138 in small municipalities, 92 in big cities)

- **Sample (gross)**: 20501 addresses

  Adresses from **population registers of the sampled municipalities** (sampling done by survey company **infas**)

- **Target sample (net)**: 4000 households

- **Realized sample (net)**: **3565 households** with 7084 persons aged 16+ (423 of these have not been (proxy-)interviewed), 8135 persons total.

- **AAPOR II Response Rate: 18.6%**

Fieldwork

Panel on household finances

- **Contact strategy (letters, telephone calls, interviewers‘ visiting)** in cooperation with **infas** (survey company)
- **CAPI Interviews** coordinated / conducted by **infas**
- **Incentive: 10 euro coin** “50 Jahre Deutsche Bundesbank”

Phase 1: Sep 14th 2010 to Feb 28th 2011

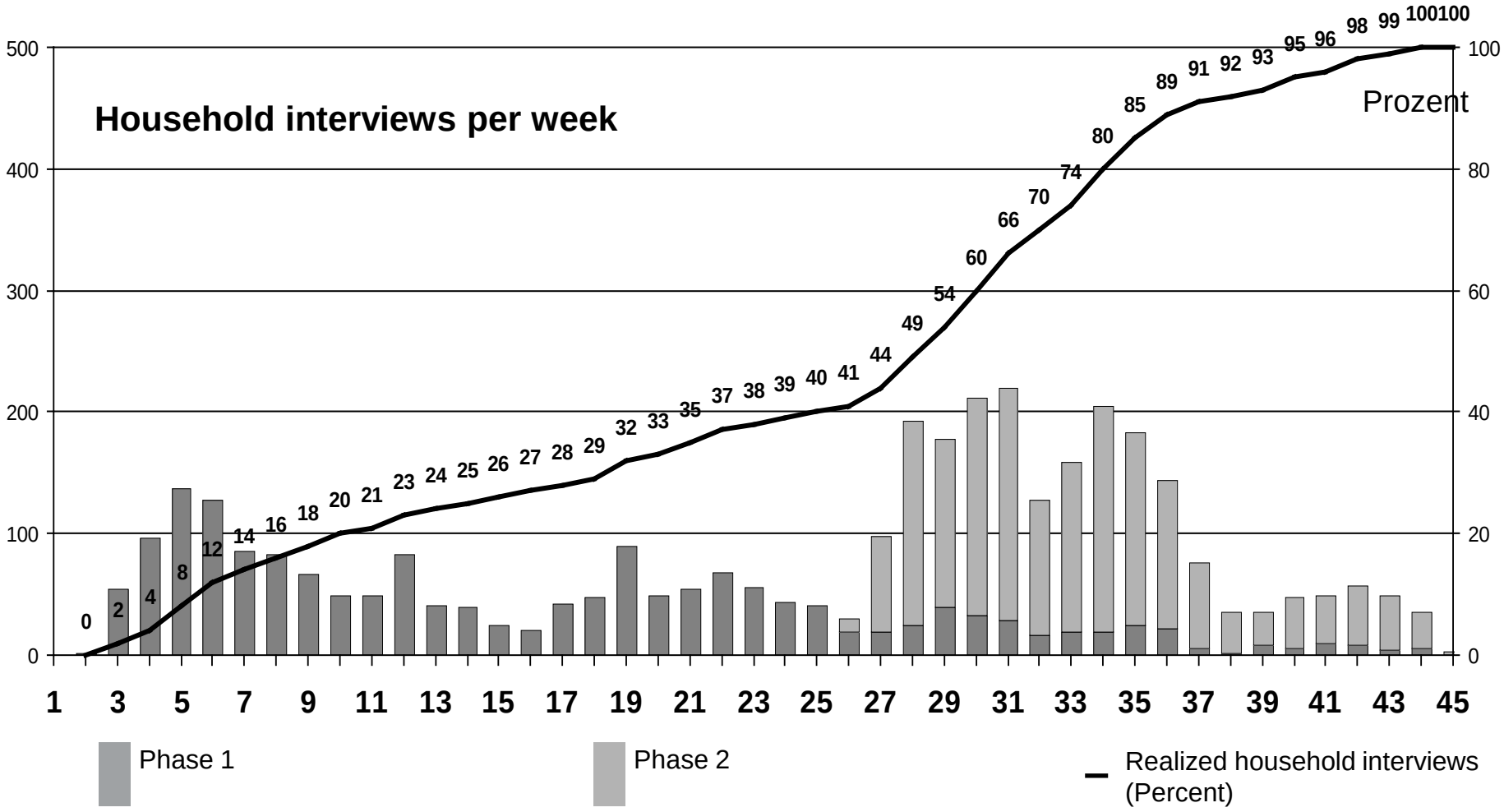
(+ Conversion of „soft deniers“ until the end of Phase 2)

Phase 2: Mar 1st 2011 to Jul 15th 2011

- Without low-performing interviewers
- Higher incentives for low-income households
- Less ex-ante information for the households
- Better Monitoring and more incentives for the interviewers

Fieldwork – Phase 1 & Phase 2

Panel on household finances



Data Quality

Panel on household finances

- **High unit non-response rate**
- **Low item non-response rates**
- **Socio-demographic structure of sample** (using design weights) corresponds very well with Microcensus information
- Totals for **real assets and mortgages** very close to aggregate statistics (e.g. national and financial accounts data, banking statistics)
- Expected difference to aggregate for **financial assets and consumer credit**

(compared to financial accounts data, banking statistics, securities deposit statistics)

Panel structure

Panel on household finances

- Planned **survey frequency**: 3 years
- Planned next field phase: spring 2014
- Target sample (net): 5000 households (ca. 3000 panel households)
- SOEP strategy: **Full panel** -- all households recontacted, all individuals tracked
- **PSID** concept: Follow all household splittings if one of the household members belongs to the original sample household
- Panel mortality: **Refreshment samples** at regular or irregular intervals

Weighting

Panel on household finances

Design weights:

- Reflect the sampling probability for every household,
- Correct for oversampling and other unequal probability sampling bias

Non-response weights:

- Correct for non-response bias (some types of households show lower / higher willingness to take part in the survey than others)

Calibrated Weights

- Calibrations (redressment) to adjust the distribution to Microcensus distribution (using sociodemographics and regional information)

Result: Final weights

Editing

Panel on household finances

Detecting and removing of inconsistencies
Improving data quality
Preparation for and interaction with imputation

} **without distorting the data**

- Initial **filter and value checks**: Mechanical data checks of the correct filtering and routing into the correct answering path
- **Structural edits**: **Recoding** verbatim answers, converting currencies into euro etc.
- **Logical consistency checks**: Testing the consistency of households' answers with regard to other answers given during the interview
- **Outlier checks** that detect whether some values for a given household were clearly too high or low in comparison to the other answers given by the same household and other households in the data set.

Edited values are marked by editing **flags**.

Imputation I (general)

Panel on household finances

Item non-response -> Missing data -> Diminishing and distortion of information

Assumption: Non-response mechanism is **„missing at random“ (MAR)**: The probability of a missing observation can be fully explained using the observed values in the data set

Multiple Imputation: Takes into account the uncertainty of the selected imputation model to avoid underestimation of variances and covariances in the imputed data set. Every missing value is replaced by a number of independently imputed values, known as **implicates**. This routine is based on the bootstrap procedure.

Iterative process until convergence is achieved.

Imputation II (PHF specific)

Panel on household finances

Imputation methods:

- Core procedure: **SAS program FRITZ**, developed by Arthur Kennickell (FED)
- **Continuous variables**: Linear stochastic regression model: Missing values are substituted by their best linear predicted value, plus a normally distributed random variable.
- **Binary variables**: Linear stochastic regression
- **Categorical variables**: Hot deck procedure

Final imputation: 14 iterations, 5 different implicates

Imputed values are marked by imputation **flags**.

Anonymisation

Panel on household finances

Starting point: Formally anonymised data (without names and addresses) delivered by infas.

Aim: Scientific Use File (SUF) satisfying established standards for SUFs and demands of Deutsche Bundesbank legal department.

Applied anonymisation methods:

- **Deletion of variables** (e.g. verbatim answers, paradata collected by the interviewer)
- **Coarsening / recoding of variables** (e.g. regional indicators)
- **Topcoding** (e.g. age – at 90)
- **Adding stochastic noise** (age – at 70+)
- **Random rounding** (most continuous variables – at two significant digits)

Questionnaire sections

Panel on household finances

1. **Screener and Demographics**
2. **Consumption**
3. **Real assets and their financing**
4. **Other liabilities / credit constraints**
5. **Participating interests (businesses, fin. assets)**
6. **Intergenerational transfers / gifts**
7. **Employment**
8. **Pensions and insurance policies**
9. **Income**

Answered by FKP (Financial knowledgable person)

Answered by every household member aged 16+ (proxy interview possible)

SUF: Data Files

Panel on household finances

- **M-file:**
  - Basic demographic information about **all** household members
  - "Household matrix": relationship between household members
- **P-file:**
  - Variables on individual level (every household member **aged 16+**).
  - If missing=1 then sections 7 (employment), 8 (pensions and insurance policies) and 9 (income) are completely imputed.
- **H-file:**
  - Variables that have been collected on the household level by interviewing the FKP
- **W-file:**
  - Replicate weights for variance estimation (on the household level)

Data structure: Basic facts

Panel on household finances

Number of variables: ca. 2200 (H,P,M) (including flags)

ID variables:

- **caseid:** Indicating the originally sampled household (does not change over waves)
- **hhid**: Household ID (can change over waves)
- **pid**: person level ID, number for household members, pid=1 for FKP
  (in some cases the pid is not consecutive because some household members dropped out of the survey)
- **Impid**: implicate ID (linked to the 5 multiple imputation implicates)

It will be possible to track specific households and persons over waves.

Variables usually have a corresponding **flag variable** indicating the status of the value.

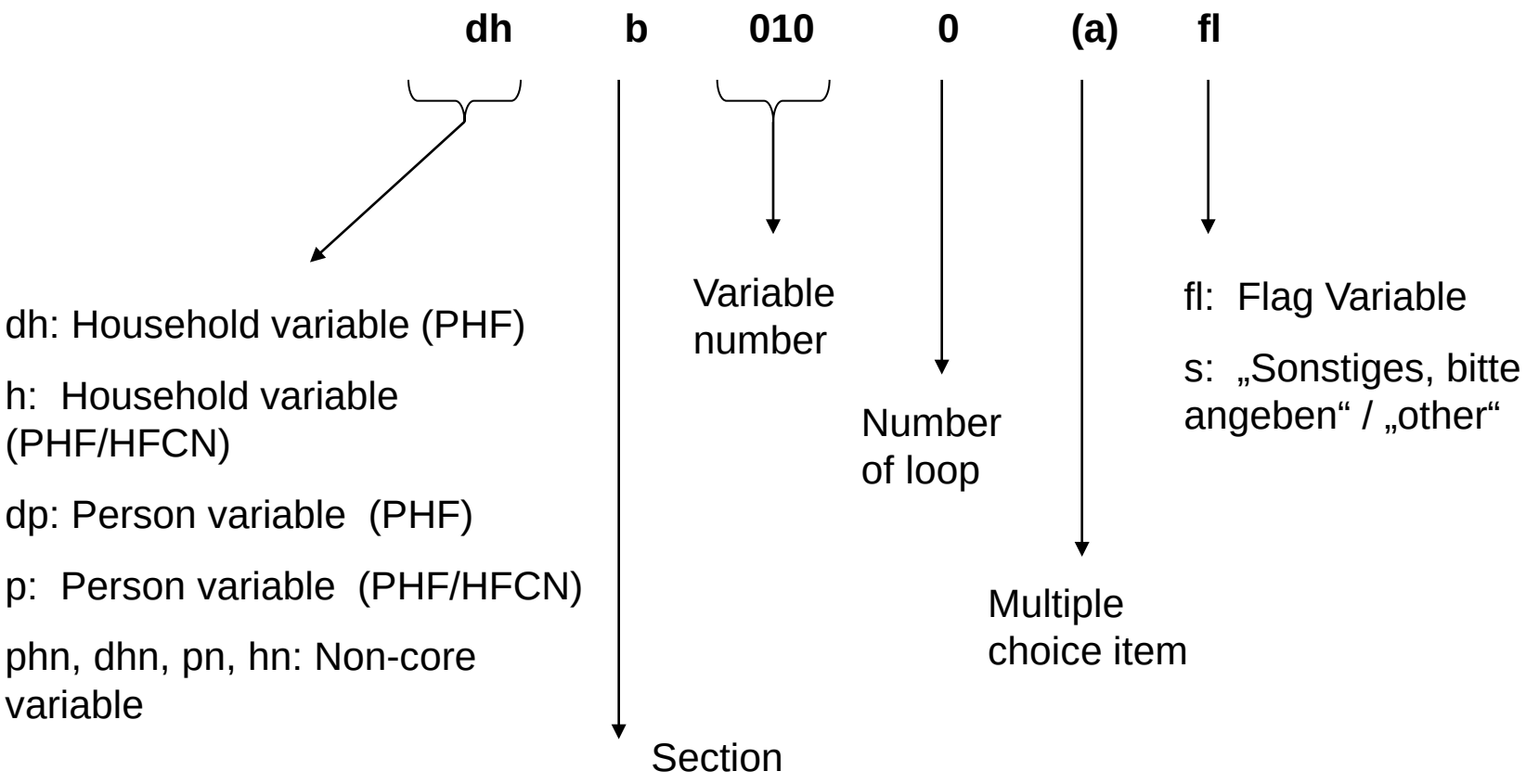
Data structure: Types of variables

Panel on household finances

- **Binary: 1, 2 (e.g. 1=yes, 2=no)**
- **Categorical: 1, 2, 3, …. (coded value label)**
- **Continuous (e.g. euro values)**
- **Missing-Codes:**
  - -1 Do not know
  - -2 Not specified
  - -3 Question filtered
  - -4, -5, -6 special codes for some few questions

Data structure: Variable names I

Panel on household finances



Data structure: Variable names II

Panel on household finances

- **vsqm*:** variables of the household matrix (relationships)
- **r* , dr*:** basic demographic variables
- **other variables** (IDs, regional indicators, weights, interview metadata etc.)

(Almost) All variables have english **variable labels**

(Almost) All Variables have english **value labels**

Flag variables

Panel on household finances

- **1-digit flags:**
  - 1 recorded as collected
  - 0 not applicable

- **4-digit flags:**
  - **1st digit**: 1=not imputed, 2=imputed
  - **2nd digit**: 0=not edited, 1=manually edited, 2=manually set to missing, 5=automatically edited, 6=automatically set to missing
  - **4th digit**: 0=original value „don't know“, 1=original value „no answer“, 2=missing value due to a missing answer in a preceding question, 3=implausible value, 4=value had been provided in interval ranges, 5=CAPI-error or interviewer-error, 6=recoded value, 7=currency conversion, 8=net-gross-conversion.

Sample design variables

Panel on household finances

stratum: three first-stage strata:

- other small municipalities,
- wealthy small municipalities,
- large cities (more than 100000 inhabitants).

stich: the second stage of sampling in large cities: Large cities are split into wealthy street sections and other street sections.

p_nr: The **PSUs** (sample points): Households having the same p_nr are located nearby. Besides this, the value of p_nr does not contain any regional information.

Weights

Panel on household finances

In the H-file:

- **exw_hh:**

  Final (calibrated and non-response-adjusted) household weights: Mean=1.

- **exhoch_hh:**

  The corresponding expansion factor: number of households in Germany that are represented by the individual household (min.: 2 ; max.: 120000)

The W-file contains 1000 replicate weights for variance estimation.

Regional information

Panel on household finances

Bundesland (bland) is recoded into four categories:

- **North** (Niedersachsen, Schleswig-Holstein, Hamburg, Bremen)
- **West** (Nordrhein-Westfalen, Rheinland-Pfalz, Saarland)
- **South** (Bayern, Baden-Württemberg, Bayern)
- **East** (Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt, Sachsen, Thüringen, Berlin)

The municipality size class (polgk) is recoded into five categories (gradations at 5000, 20000, 100000 and 500000 inhabitants).

The BIK region size class (variable bikgk) is recoded into five categories (originally ten BIK categories): <50000; 50000-500000 type 2/3/4; 50000-500000 type 1; >500000 type 2/3/4; >500000 type 1.

Due to further anonymisation, polgk and bikgk are missing for wealthy small municipalities in the Region East.

ISCO, NACE

Panel on household finances

Occupation:

- Coding following International Standard Classification of Occupation **ISCO 88**,
- mainly two digits,
- special code: **79** = manufacturing: supervisor / foreman

Branch of economic activity:

- Coding following **NACE Rev. 2 from 2008**:
- **A-U -> 1-21** in PHF

Conversion: time period, income, interest rates

Panel on household finances

Caution! There are some **differences between questionnaire and data set variables**:

In the questionnaire there are questions (e.g. dhc0510; the sequence dpg0300, dpg0310, dpg0320) asking for the **time period** of values (monthly, quaterly, yearly). In the data set these variables are missing, because all variables with time period information contain **yearly values**.

The interviewed household members had the possibility to provide gross or net income values. Before imputation all net values had been converted into **gross values**.

The questionnaire variables related to nominal interest rates, e.g. dhb5611, are missing in the SUF (contains only **effective interest rates**).

Comparability with HFCS data

Panel on household finances

- **You can match data records via household id: sa0010 = caseid (= hhid)**
- **Variables named p... or h... are matching 1:1 following the HFCN definition**
- **In general, the comparability is high, but be careful!**
  - Some question texts are similar, but not identical.
  - The questionnaire structure differs at some points.
  - Some HFCS variables (and flags) have been calculated ex post by combining several PHF variables (flags).

Documentation, reports

Panel on household finances

- **Readme-file** (basic information about data structure) is disseminated together with the PHF-SUF

- **Bundesbank Discussion paper 13/2012**: Ulf von Kalckreuth, Martin Eisele, Julia Le Blanc, Tobias Schmidt, Junyi Zhu:

  The PHF: a comprehensive panel survey on household finances and wealth in Germany

- **Bundesbank monthly report Jan 2012**: The PHF: a survey of household wealth and finances in Germany

- **To appear: Volume** containing several documentation, methodological (and other) reports

Data application

Panel on household finances

http://www.bundesbank.de/phf-data

phf-data@bundesbank.de

Only for research purposes – commercial use of PHF data is strictly excluded!

We need (from everyone who uses the data):

1.) Completed research data request form (by email)

2.) Copy of identity card or passport

3.) Curriculum vitae (by email)

4.) Completed and signed certification document (by mail)

Encryption: http://www.7-zip.org/

Data dissemination

Panel on household finances

After your application has been confirmed:

-> encryptet PHF-data sent by email:

1.) H-file, P-file, M-file (Stata; 2.5 MB)

2.) W-file (Stata; 18 MB !)

-> password transmitted per telephone

Please use your Project-ID in the email-communication!

Feedback

Panel on household finances

Feel free to give us feedback

-> improvement of data quality

-> improvement of documentation

-> improvement of data application procedure

Contact:

phf@bundesbank.de
martin.eisele@bundesbank.de
junyi.zhu@bundesbank.de

DEUTSCHE BUNDESBANK
EUROSYSTEM

Thank you for your attention !