Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Distributional Effects of Inflation in the Czech Republic

Author: **Bc. Petra Linhartová**

Supervisor: **Petr Janský, Ph.D.**

Academic Year: **2014/2015**

Bibliographic Evidence

LINHARTOVÁ, Petra: *Distributional Effects of Inflation in the Czech Republic*. Prague, 2015. 93 pp. Master's Thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis Supervisor: Petr Janský, Ph.D.

Number of characters: 113,849

Declaration of Authorship

I hereby declare that this thesis has been compiled independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

I grant to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 12, 2015

Signature

Acknowledgments

My gratitute belongs especially to my supervisor Petr Janský for providing insightful recommendations and helping me throughout the writing of the thesis. I am grateful to Pavel Hait for the encouragement, valuable remarks and suggestions leading to further improvement of this thesis.

I would also like to thank to my friends, students of the Institute of Economic Studies, for their support and cooperation during our studies and during writing of the thesis. Special thanks belongs to Andrea Yeash and Carl A. Svoboda for English proofreading.

Abstract

Consumer price index captures the changing costs of the consumer basket of a typical household. Despite differences in spending patterns, change in consumer price index is used as a measure of inflation for the whole population. The aim of this thesis is to assess how close to the official inflation rate households are and determine which groups have significantly different inflation. Using the Czech data from the Household Budget Survey over the 1990–2012 period we calculated specific inflation for each household in our sample. We first found out that on average only two thirds of households are close to the official inflation rate, which led us to the construction of subgroup price indices. In the empirical part, we examined the effect of household characteristics on inflation by applying the fixed effects estimation. We found that low-income households, pensioners, households in urban areas and households with few members have higher than average inflation.

JEL Classification	C33, C42, C54, E31, R29
Keywords	inflation, inflation differentials, households, average inflation representativeness, subgroup inflation rates, fixed effects

Author's e-mail	p.linhartova@gmail.com
Supervisor's e-mail	jansky@fsv.cuni.cz

Abstrakt

Index spotřebitelských cen zachycuje měnící se náklady spotřebního koše průměrné domácnosti. Změna indexu spotřebitelských cen vyjadřuje míru inflace pro celou populaci, přestože výdaje domácností jsou různé. Cílem této diplomové práce je zhodnotit, jak blízko jsou domácnosti oficiální míře inflace a určit, které skupiny obyvatel mají významně odlišnou inflaci. Za použití českých dat ze Statistiky rodinných účtů pro období 1990–2012 jsme vypočítali specifickou inflaci pro každou domácnost v našem vzorku. Nejprve jsme zjistili, že v průměru mají pouze dvě třetiny domácností míru inflace blízkou oficiální míře inflace, což nás vedlo k sestavení cenových indexů pro různé skupiny obyvatel. V empirické části jsme se zabývali vlivem charakteristik domácností na inflaci za použití panelové metody fixních efektů. Ukázalo se, že domácnosti s nízkými příjmy, penzisté, domácnosti ve městech a domácnosti s méně členy mají vyšší než průměrnou inflaci.

Klasifikace JEL	C33, C42, C54, E31, R29
Klíčová slova	inflace, inflační rozdíly, domácnosti, reprezentativnost průměrné inflace, míra inflace skupin obyvatel, fixní efekty

E-mail autora	`p.linhartova@gmail.com`
E-mail vedoucího	`jansky@fsv.cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

CNB	Czech Nationa Bank
COFOG	Classification of the Functions of Government
COICOP	Classification of Individual Consumption by Purpose
COPNI	Classification of the Purposes of Non-Profit Institutions
CPI	Consumer Price Index
CZSO	Czech Statistical Office
EU-SILC	European Union - Statistics on Income and Living Conditions
HBS	Household Budget Survey
FE	Fixed Effects estimator
FEI	Fixed Exact Index
MAD	Mean Absolute Deviation
OLS	Ordinary Least Squares
OECD	Organisation for Economic Cooperation and Development
PCEPI	Personal Consumtion Expenditures Price Index
PPI	Producer Price Index
RE	Random Effects estimator
RPI	Retail Price Index
SD	Standard deviation
UNSD	United Nations Statistics Division
UZIS	Institute of Health Information and Statistics of the Czech Republic

Master's Thesis Proposal

Author	Bc. Petra Linhartová
Supervisor	Petr Janský, Ph.D.
Proposed topic	Distributional Effects of Inflation in the Czech Republic

Motivation The main measure of the inflation in the Czech Republic is Consumer Price Index (CPI), the cost of goods and services purchased by typical household. But is there such a thing as a typical household? We may consider CPI to be a good measure of inflation only if it holds for everyone. The question of representativeness of the inflation rate has not often been addressed: Crawford et al. (2002) or Leicester et al. (2008) examine whether the inflation in the UK is representative for any household. I will build upon Hait and Janský (2014) and verify the representativeness of inflation in the Czech Republic over the last 25 years.

If the inflation differs significantly across Czech households, it is interesting to search for types of households suffer the most from inflation. Results throughout existing literature differ quite significantly, in the case of low-income households some studies conclude that they face higher than average inflation, others that it is equal to the average or lower. I will analyse the case of the Czech Republic not only for low-income households, but also high-income, young households and pensioners. Inflation components, which are published quarterly by CNB in Inflation reports, will enable me to see whether the decomposition of inflation differs from the micro and macroeconomic point of view. Besides that I will, similarly to Ball et al. (2013) who did the analysis for the fiscal policy, examine the distributional impact of the change of inflation on inequality; impulse response functions will be used for it.

In the last part of my thesis I will examine the effects of monetary policy shocks on income inequality, which seems to play a bigger role in driving inequality patterns. Coibon et al. (2012) have proved with the US data that con-

tractionary monetary shocks significantly raise the observed inequality. Such quantification may be the first step to identify what kind of policy would be most suitable for the Czech Republic.

Hypotheses

1. Pensioners and low-income households face higher inflation than average.
2. High inflation increases inequality between households.
3. Macro and microeconomic decomposition of inflation are equivalent.
4. Inequality increases after monetary policy shocks.

Methodology I will use the data that CZSO uses for the construction of consumer price indices Household Budget Survey (detailed data on expenditures, income and other characteristics for 3,000 Czech households). I will examine the representativeness of inflation, its distributional impact and inflation for different subgroups: young households, pensioners, low- and high-income households. This part of my thesis will prolong Hait and Janský (2014) analysis for the period of 1989–2013, add more details and an analysis of the distributional impact of inflation. Inflation reports published quarterly by Czech National Bank will help us connect the micro and macroeconomic point of view, compare it and see whether specific reasons of price increases do lead to higher inequality of different groups.

In the second part of the thesis, I will analyse the effects of monetary policy shocks on income inequality. Similarly to Coibon et al. (2013) I will use the VAR analysis to see whether monetary shocks have an impact on inequality, the direction of the potential effect and its extent.

Outline

1. Introduction
   (a) Motivation behind the thesis
   (b) Hypotheses
2. Literature overview
   (a) Inflation differentials and distributional impacts of inflation
   (b) Monetary policy and inequality
3. Inflation - empirical part
   (a) Representativeness of average inflation rate

   (b) Specific inflation rates for different types of households
   (c) Distributional impact of inflation
   (d) Decomposition of contributors to inflation
   (e) Macroeconomic comparison
4. Monetary policy - empirical part
   (a) Identification of monetary policy shocks
   (b) Effects of monetary policy on inequality - VAR analysis
5. Conclusion and discussion of the results

Expected contribution The research done on Distributional Impact of Fiscal policy is much broader than in case of Monetary policy and inflation, there are no papers describing the situation in the Czech Republic according to my knowledge. I will try to fill this gap and to find out what are specific price increases and monetary policy tools that increase the inequality the most. Due to very long time period (1989–2013) I will examine the whole transformation period until now to estimate overall distributional effect of inflation over the time. The biggest contribution of this work is combination of long-term detailed data examined and econometric approach of VAR analysis.

Note During the work on the thesis, the topic was narrowed and also the title changed. Instead of focusing on the whole monetary policy, we analysed only inflation and VAR analysis was not performed, instead we conducted fixed effects estimation of panel data. Nevertheless, this thesis proposal can serve as a springboard for further research.

Core bibliography

1. Artsev, Y. ; V. Roshal; & Y. Finkell (2006): "Consumer Price Indices - Measuring Across Households." *Ninth Ottawa Group Meeting on Prices*, London.
2. Ball, L. ; D. Furceril ; D. Leighl; & P. Loungani (2013): "The Distributional Effects of Fiscal Consolidation." *IMF*. Working Paper 13/151.
3. Coiboni, O. ; Y. Gorodnichenko ; L. Kueng; & J. Silvia (2012): "Innocent Bystanders? Monetary Policy and Inequality in the U.S.." *IMF*. Working Paper 12/199.
4. Crawford, I. ; & Z. Smith (2002): "Distributional aspects of inflation." *Institute for Fiscal Studies*. Commentary 90.

5. Garvey, E. ; & E. Murphy (2004): "A consumer price index for low-income households in Ireland (1989-2001)." *Combat Poverty Agency.*

6. Hait, P. ; & P. Janský (2014): "Inflation differentials among Czech households.." *CERGE-EI. Working paper series.* ISBN 978-80-7344-305-4.

7. Hobijn, B. ; & D. Lagakos (2003): " Inflation inequality in the United States." *Review of Income and Wealth* 51(4), 581 - 606.

8. Leicester, A. ; C. O'Dea ; & Z. Oldfield (2008): "The Inflation Experience of Older Households." *Institute for Fiscal Studies.* Commentary 106.

9. Lommatzch, K. (2009): "Monetary policy and the inflation process prior to EMU membership: The Czech and Hungarian experience.." *University of Berlin.* Academic dissertation.

10. Romer, C. D. ; D. H. Romer (1998): "Monetary Policy and the Well-Being of the Poor." *National Bureau of Economic Research.* Working Paper No. 679.

11. Stewart, K. J. (2008): "Experimental consumer price index for elderly americans (cpi-e): 1982-2007." *Monthly Lab. Rev.* 131, 19.

12. Sugema, I. ; M. I. Irfany ; A. Holis; & T. Bakhitiar (2010): "Consumer Price Index for the Poor (CPI-P): An Empirical Analysis of Indonesia." *International Research Journal of Finance and Economics.* Issue 58.

Author

Supervisor

Chapter 1

Introduction

The main measure of the aggregate inflation in the Czech Republic is the Consumer Price Index (CPI), which is a measure of the changing costs of goods and services purchased by a typical household. But is there such thing as typical household? Can we believe that a pensioner has the same consumer basket as a young family? This thesis aims to answer these questions.

Even though the CPI may provide good measure of the average inflation rate, it would be surprising if it was a good measure for everybody. As prices of goods and services rise at different rates, households might have different inflation rate due to variation in the content of their consumer baskets. Differences in inflation rates are visible from the studies conducted in other countries and Czech Statistical Office publishes consumer price indices for the whole population, for pensioners only and for households living in Prague and these indices are significantly different. One of the goals of this thesis is to quantify these differences to be able to evaluate whether they are significant and constant or not. If most households are close to the average inflation rate, we may say that consumer price index is a good proxy for the whole population, however, if the opposite is true, the use of specific subgroups price indices should be considered. To assess the "closeness", we calculate specific inflation rates for every household in our sample and construct three measures to examine how well consumer price index represents the experience of all households. We also introduce a new measure of closeness, which is suitable for years with high, but also with a low level of inflation.

The inspiration for this thesis are mainly studies of Crawford & Smith (2002), Leicester et al. (2008) and Hait & Janský (2014) that investigated not only how close the average inflation rate is to the experience of the majority of

households, but also constructed subgroup inflation rates for groups of similar households that experience specific inflation rate. This thesis builds on the work of the aforementioned authors and further expands their research.

Similarly to these studies we construct subgroup inflation rates for selected groups that might have similar spending patterns. The aim of this qualitative part is the examination of the extent to which inflation varies within a particular household group. Our selected groups are low- and high-income households, pensioners and single person households, households with children and households in rural and urban areas.

The main value added of the thesis is the econometric analysis performed in Chapter 6. It allows us to analyse all the effects together and assess the impact of a variable on inflation holding other effects fixed. If subgroup inflation rates are constructed and pensioners are found to have higher inflation, we must take into account that the reason behind might be different (e.g. all pensioners might have low income). Panel data analysis helps us to exclude the effects of other variables and quantify their effect. As to our knowledge, no other study on this topic has conducted such analysis. We first discuss the data, drawbacks of survey panel data and choice of the estimator. After the robustness testing we perform fixed effects estimation with time effects and clustered standards errors. The last section presents the results and proposes areas for further research.

The structure of the thesis is as follows: next Chapter 2 is the review of existing literature on the representativeness of the inflation rate and on construction of subgroup inflation rates. Chapter 3 describes the data used in this study and the process of their adjustment. Chapter 4 is dedicated to the average inflation rate representativeness, we introduce three measures of the "closeness" to the average inflation rate and methods of calculation of household and subgroup inflation rates. The empirical part is divided into two chapters: qualitative part in Chapter 5 constructs subgroup inflation rates for selected groups and quantitative part in Chapter 6 examine the effects of microeconomic characteristics of households on inflation. Finally, a conclusion of the whole thesis follows in Chapter 7, Appendix A lists COICOP classes and Appendix B contains tables with results.

Chapter 2

Literature review

This paper addresses two topics related to unequal inflation of households - the average inflation rate representativeness and construction of subgroup inflation rates with related empirical estimation of the effects of household characteristics on inflation. The literature review relevant for this work is also divided into two separate subsections and follows in Section 2.1 and in Section 2.2.

2.1 Consumer Price Index (CPI)

Consumer Price Index (CPI) as a measure of inflation has two main drawbacks as discussed further in Subsection 4.2.1. Many studies focused on the first drawback - biases that are linked to fixed-based price index. The authors either try to estimate the magnitude of CPI bias or try to compute specific true cost-of-living indices. The second drawback, the question of representativeness of the average inflation rate, has received less attention. CPI bias moves the inflation for the whole society and policy makers may account for it, but if everyone has different inflation, it becomes hard to deal with it.

2.1.1 CPI bias

The results of CPI bias worldwide differ only in magnitude, all find a positive CPI bias, which ranges from 0.3 to 1.1 percentage points. Leszczynska & Halka (2012) summarized in their paper the results of many studies and review from Neves & Sarmento (1997) is replicated in Subsection 4.2.1. (for more details see Table 4.1)

We are more concerned about the empirical evidence for the Czech Republic. Hanousek & Filer (2002) examined the CPI bias with the use of Fisher Exact Index (FEI) during the transition period for consumer homothetic behaviour.[1] According to their results, CPI overstated the FEI by 0.1 to 0.8 percentage points during the period 1993–1999. Podpiera (2003) abolished the assumption of homotheticity and his results for 1994–2000 showed that the yearly CPI bias ranged from -0.83 to 0.51 percentage points. He also highlighted that in periods when the inflation was around 2%, the absolute bias tends to stabilize at a level of 0.5 of a percentage point in absolute value. This result was relevant for years after 2000 when the inflation was less volatile than in the 90s.

2.1.2 Representativeness of the average inflation rate

The question of representativeness of inflation has not been addressed many times: Crawford & Smith (2002) or Leicester et al. (2008) examined whether the inflation rate in the UK is representative for all households and Artsev et al. (2006) applied the same methodology on Israeli data. There is a consensus that the dispersion in rates among individual households during times of high average inflation rates tends to be higher than during period of low inflation.

Crawford & Smith (2002) used the data from 1976 and 2000 to investigate the distribution of inflation rates and the differences between various group price indices. The households that were within one percentage point from the official inflation rate ranged from 10 to 70%. Leicester et al. (2008) came to the same conclusion on the UK data prolonged to 2008 and pointed out that households tend to be further from the official inflation when the inflation is high.

Artsev et al. (2006) showed on the data from 1999 to 2005 that 52–65% of households were within one percentage point from the mean rate of inflation and concluded that the overall consumer price index is generally a good reflection of the inflation for most households.

As to my knowledge, the only study that examined representativeness of Czech average inflation rate was the one by Hait & Janský (2014). They have

[1] Hanousek & Filer (2002) defined the Fisher Exact Index (FEI) as "the geometric mean of a Paasche and a Laspayres Index, is what is generally referred to as a "superlative index" that, under reasonable assumptions, is a true cost of living index in that it measures the change in incomes that would be required to keep a consumer utility neutral following a change in prices."

found that on average 64% of households were within one percentage point from the official inflation rate in 1995–2010. The dispersion was lower than in case of the UK results - the peak was 87% in 2003 and the lowest representativeness was in 2008, when only 42% of households had the inflation rate within one percentage point of the average.

The existing literature on the topic shows that it is relevant to address the question of the appropriateness of inflation rate measures. This paper will prolong the examined period and study the whole transformation since 1990, add the most recent data for 2011 and 2012 and compare the results with the existing studies mainly from the UK.

2.2 Subgroup price indices

If we conclude that the inflation is not the same for every household, it is reasonable to construct price indices for households that might have similar spending patterns and therefore we expect them to have the same level of inflation. The use of subgroup price indices in policy making is questionable, but the differences should be monitored on a regular basis.

The history of subgroup price indices goes back to 1958, when Arrow (1958) noted that people with low income have different spending patterns than high income groups as they spend more of their budget on necessities. The work done in this field can be generally divided into two parts according to definition of groups. The groups are defined either according to income (studies focusing on relatively poor households or comparison of low and high income groups) or according to specific household characteristics (age, family structure or regional differences). Most of the studies came to the consensus that the difference in the rates of inflation is not very high and if there is a bigger difference, it is not valid in the long run.

Subgroup price indices - income

Garner et al. (1996) constructed experimental price indices for the US poor and compared it with the whole population in 1984–1994. They found very little difference between the two price indices. They also constructed Fisher index for the poor and it was more or less the same as the index for the whole urban sample.

Murphy & Garvey (2004) analysed urban and rural price indices for the

lowest income groups on the Irish data between 1989–2001. They found that all the indices were similar until 1996, but since then prices grew more for the urban poor group. The main reasons behind this increase were rental costs, mortgage interest, cigarettes and clothing.

Oosthuizen (2007) used the data for South Africa to conclude that neither rich nor poor households have experienced rates of inflation consistently higher or lower than average in 1998–2006.

Sugema et al. (2010) showed that there was a significant difference between CPI for the poor (particularly in rural areas) and CPI of overall population in Indonesia from 2006 until 2009.

Subgroup price indices - household characteristics

Idson & Miller (1999) found that US families with children experienced relatively lower inflation rates than families without children during the 1968–1987 period. They concluded that families with children are usually younger, and younger families have lower rates of inflation. This was confirmed by many studies focusing on elderly households, which tend to have higher than average inflation.

Crawford & Smith (2002) investigated besides representativeness of the average inflation rate also price indices for different groups. They found out that over the whole period examined, non-pensioners, mortgagors, employed and childless households faced higher than average inflation.

Hobijn & Lagakos (2005) used the data on US households for the period 1987–2001 to find that inflation is usually higher for elderly (mainly because of rising prices of health care). Surprisingly, they found out that households facing high inflation one year; do not generally face higher inflation the next year, the household-specific persistence in inflation disparities is low.

Artsev et al. (2006) showed that the inflation rates for different subgroups in Israel during 1999–2005 differed and they confirmed the findings of Hobijn & Lagakos (2005) that groups that experienced higher inflation one year did not face high inflation in the next year. Therefore they did not find persistence in inflation disparities.

Leicester et al. (2008) focused on the differences of inflation rates between pensioners and non-pensioners for 1977–2008. The differences were quite small with the exception of few years when the differences were quite substantial. Pensioners faced higher inflation in the last two years of the examined period

mainly due to increases in the price of fuel and food that are the most important part of their shopping basket.

Subgroup indices - Czech evidence

The Czech Statistical Office (CZSO) calculates overall price index, as well as two group specific price indices - for pensioners' households and for households living in Prague. As a consequence of data availability, there are many web and journal articles that compared pensioners' households with an average household as the data are publicly available. The evidence for other groups is very rare, therefore it is useful not to focus primarily on pensioners (even though they are the first group of interest that comes to mind and many studies on subgroup indices choose them as the examined group), but on other groups of households (low- and high-income households, households with many members, families with children or households in rural or urban area).

Hait & Janský (2014) constructed group specific indices for pensioners, Prague households and for a low-income sample. They have estimated that cumulative inflation from 1994 to 2010 was 209% for pensioners and 213% for low-income households. Cumulative average inflation for the whole sample in this period was only 193%, so the difference exists and it is significant. The study also pointed out that policy makers should consider whether the indexation of pensions that is now dependent on the whole sample inflation rate ("official" inflation) should not be dependent on pensioners' specific inflation rate instead.

Previous research showed that it is reasonable to construct subgroup price indices. In this paper, we will assess the extent of the difference not only for pensioners, but also for households with children or households living in a small village.

Chapter 3

Data description

3.1 Price data: Consumer Price Index (CPI)

The price data used in this work are section indices for Consumer Price Index (CPI). The data, collected by the Czech Statistical Office (CZSO), belongs among the most important indicators of inflation in the Czech Republic and they are the most disaggregate price data available for research. The methodology of construction of CPI is regularly updated. The most common changes are the adjustment of the whole consumer basket, change of weights of goods selected in the consumer basket and the choice of price representatives. Based on the methodological guidebook for use of the CPI, CZSO (2015a), consumer baskets are updated every five years, the weights are changed every two years, the choice of price representatives annually and methodology according to needs of the CZSO or demands from the European Union.

According to overall statistics from the OECD (2015), 65,000 prices are collected each month from 9,000 outlets in the Czech Republic covering around 700 selected representative items. The survey contains goods and services of the first quality currently purchased by the reference population. Both domestic products and imports are included; excluded are second hand or in any way impaired goods except for second hand cars. The CPI is further classified according to Classification of Individual Consumption by Purpose (COICOP) to 12 divisions (first level), 45 groups (second level), 105 classes (third level) and around 700 respective items (forth level).[1] This classification was developed

[1] There exists two more divisions that are not applicable to individual households; they measure individual consumption expenditures incurred by non-profit organizations and government respectively. Division 13 of COICOP corresponds to Classification of the Purposes

by United Nations Statistics Division (UNSD) in 2001 and its aim was the introduction of one comprehensive and accurate classification system to allow the comparison between countries and to avoid the double count or omission of certain sectors. General twelve divisions are presented in the Table 3.1 below, the complete list of COICOP classification can be found in Appendix A.

Table 3.1: COICOP first level divisions

Number	Name
01	Food and drinks
02	Alcohol and tobacco
03	Clothes and shoes
04	Housing, water, electricity, gas
05	Household equipment
06	Health
07	Transport
08	Communication
09	Recreation and culture
10	Education
11	Restaurants and hotels
12	Other goods and services

Source: Statistical metainformation system, Czech Statistical Office

The detailed data on price indices of individual items (forth level) are available only for internal purposes of CZSO, only the data on the third level (up to classes) are available for the research. We present price indices the period 1989-2013 in this subsection, even though period 1990–2012 is analysed throughout the thesis. Originally, we wanted to analyse the period 1989–2013, but due to data availability of Household Budget Survey, we restricted the time period. We plan to further prolong the analysis in future that is why we describe the evolution of the inflation for longer than examined period.

Unfortunately, the data for 1989 and 1990 are not available, but we were able to construct approximate indices based on the price data available. The CZSO publishes Statistical Yearbook annually, where prices of selected items can be found. Price data on 74 identical goods and services over the period 1987–1993 were gathered for the purpose of this work. Based on these prices

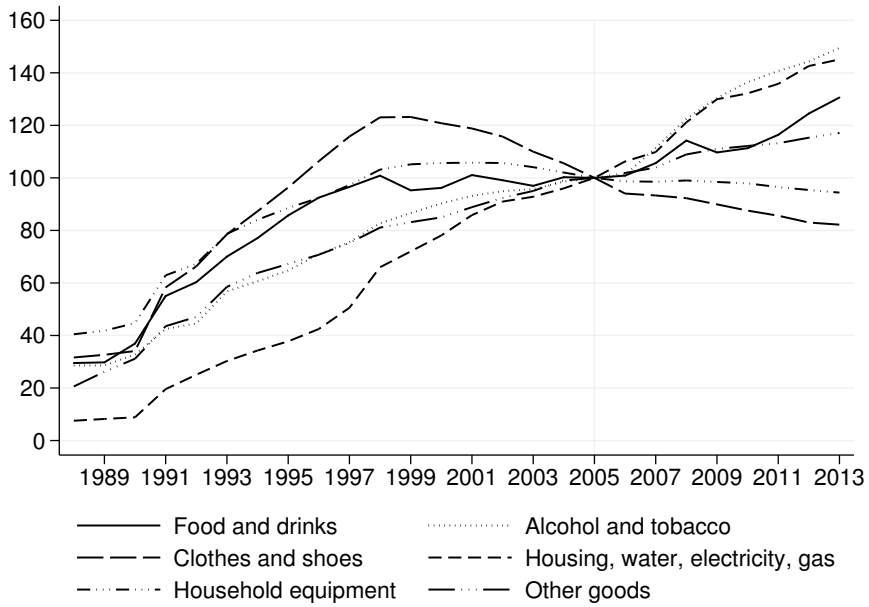
of Non-Profit Institutions Serving Households (COPNI); Division 14 of COICOP corresponds to Classification of the Functions of Government (COFOG).

and price increases, the price indices for 1989 and 1990 were calculated. Where the data were not available, the average of three consecutive years was used.

Subsequent tables illustrate the development of price indices for the period 1989–2013, where year 2005 is used as a base year. Figure 3.1 illustrates the development of prices of goods while Figure 3.3 shows the development of services prices. There is not a big difference between the two; price indices of services grew more evenly than goods. The reason behind this trend might be homogeneity - services are more comparable, the value added is mostly human work, while goods prices differ more. Both tables have the graph of inflation below to better illustrate the variance of inflation.

The price of housing, energy and alcohol grew during the whole period, but more interesting is the development of clothes prices and household equipment. It grew sharply until 1998 and from then on, it is decreasing until now. The goods inflation follows the aggregate inflation, only housing is significantly higher in 1991 and 1998 than inflation of other goods.

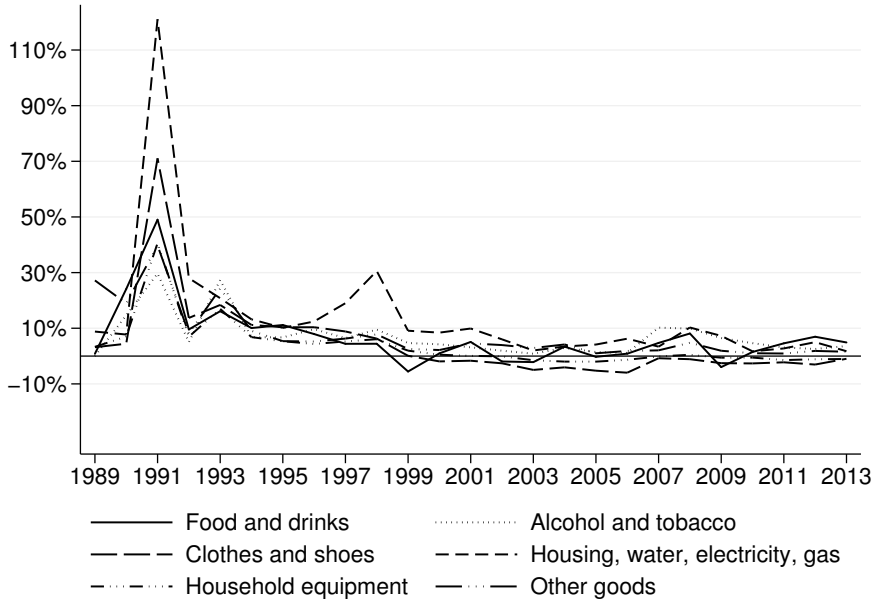
Figure 3.1: Goods price indices



Source: CZSO data and author's calculations

Figure 3.3 illustrates that services prices developed similarly except for one - price of health. The totalitarian regime had the so-called free health care and promoted it as one of the main advantages of the regime. Liberalisation of health care led to an increase in prices and since then, prices are still increasing.

Figure 3.2: Goods inflation



Source: CZSO data and author's calculations

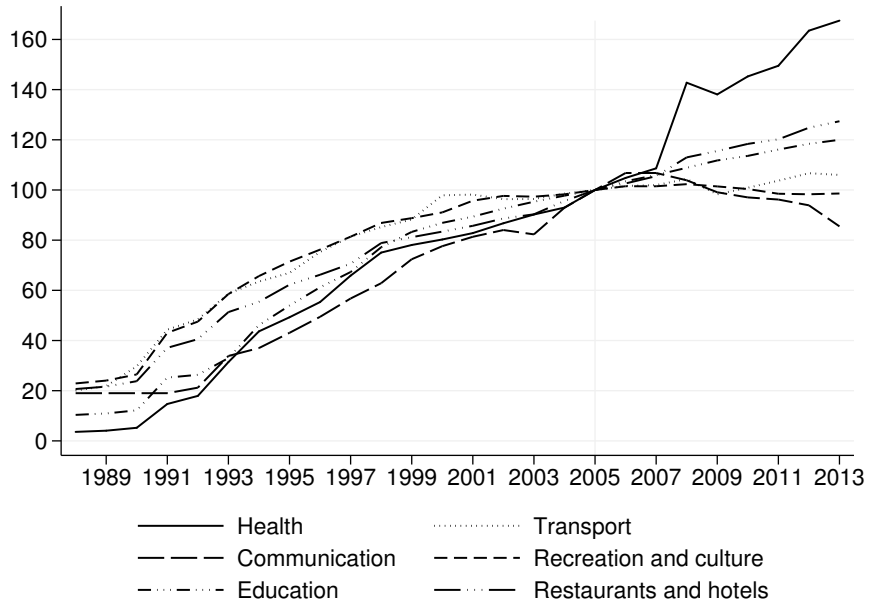
The inflation of health care in 1991 was the highest among all the divisions; it was above 180%. Another peak was in 2008 when overall inflation increased to 6.3%, while consumer prices of health increased by 26.93%.[2]

All the graphs nicely illustrate the historical development of prices in the Czech Republic after 1989. The country dealt with post-communist development of inflation relatively well, so it is useful to summarize briefly the main historical events as it will help to understand the analysis later on.

The inflation was artificially kept low by fixed prices during the time of centrally planned economy. Prices should grow annually from 1–3%, but it was not reflected in prices for the end consumers. Inflation officially did not exist, but it showed itself in other forms - lack of some goods and development of black market. Prices were kept constant and the price increase was carried out from time to time stepwise. These increases were enormous; a 50% increase was not an exception. Central bank at that time completely substituted the activity of commercial banks and was subject to the fulfillment of the state plan.

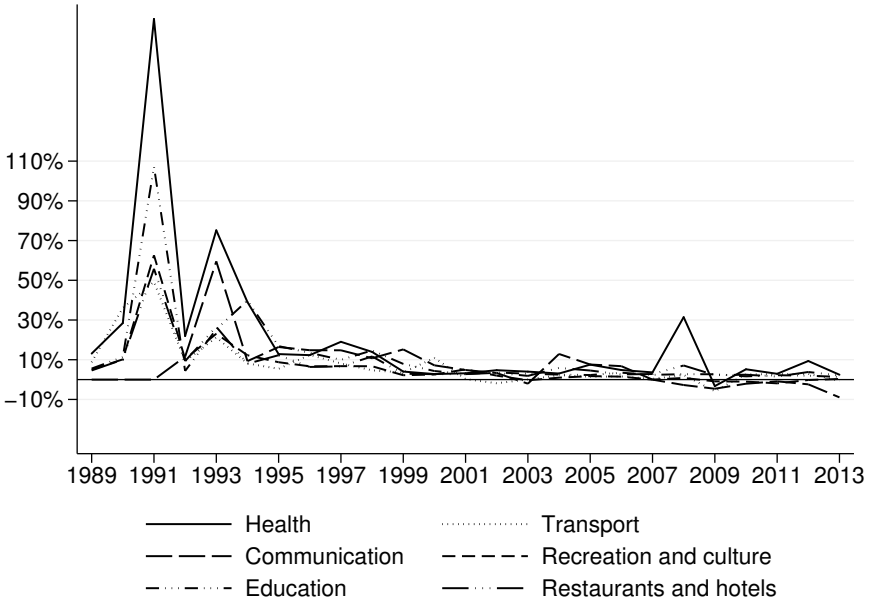
[2] For more detailed information, please refer to UZIS - Institute of Health Information and Statistics of the Czech Republic. The institute publishes every year report - Economic Information on Health.

Figure 3.3: Services price indices



Source: CZSO data and author's calculations

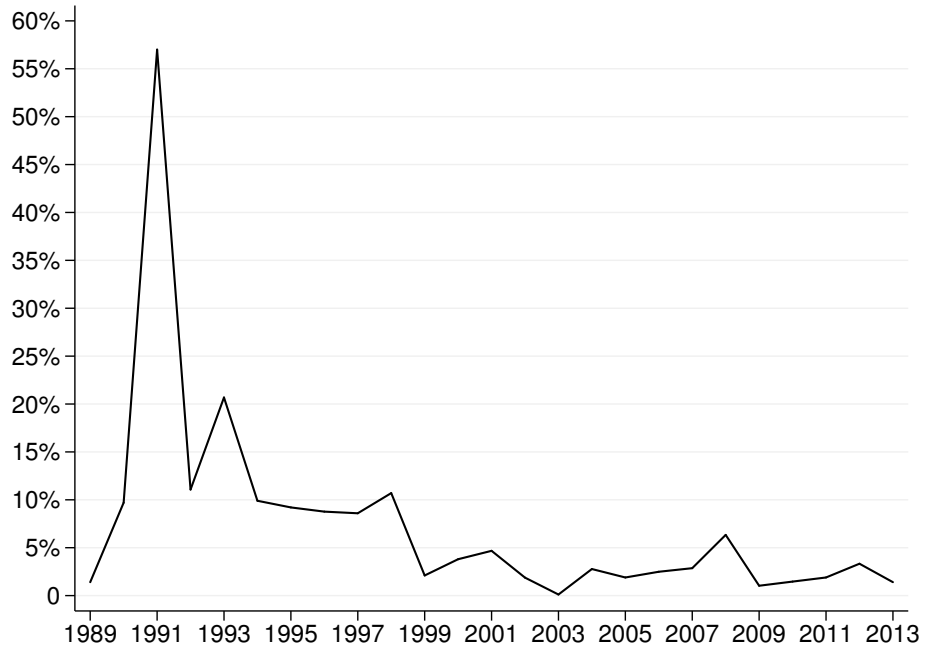
Figure 3.4: Services inflation



Source: CZSO data and author's calculations

With the introduction of market economy, prices and inflation unfreezed. The role of the national bank has changed, it became independent and its lawful goal was to maintain the stability of domestic prices and the exchange rate. The Czech National Bank (CNB) reached this goal firstly by influencing the amount of money in the economy and maintaining fixed exchange rate. In 1990, the subsidies of consumer prices were reduced and cost of fuel increased, this led to almost 10% inflation. Price liberalisation in 1991 led to the high level of inflation (56.6%), but the Czech Republic succeeded to avoid the inflation spiral and hyperinflation. Fortunately, it fell the following year to 11.1% which was the lowest figure among all post-communist countries according to Sojka (1994). The next year's introduction of value added tax in 1993 again increased the inflation to 20.8%. Fixed exchange rate regime showed to be unsustainable and the transition to inflation targeting was speed up by speculations on the Czech currency in 1997. Transition to a new regime of monetary policy did not change the goals of the CNB, it only changed the way the goals are reached. Inflation targeting regime helped to decrease the inflation in the coming years to the level of advanced economies and it is kept below 5% since then with the exception of year 2008. Figure 3.5 shows the inflation for the whole period.

Figure 3.5: Czech Inflation in the period 1989–2013



Source: CZSO data

3.2 Expenditure data: Household Budget Survey (HBS)

The data on household expenditures that we use in this study come from the Household Budget Survey data collected by the CZSO from 1990 to 2012. The HBS is an annual survey of around 3,000 households that are selected by intentional quota sampling.[3] It is important to have in mind that intentional quota sampling was used when we generalize some results for the whole population.

The dataset contains detailed information on more than 200 groups of goods and services; there is also characteristic of each household and its income. Expenditure data are obtained by method of continuous records, where every household fills the detailed diary for selected two months of the year (only total amount is reported for other months). Households are included in the dataset for the whole year if one of their key characteristics does not change (i.e. economic activity of the key person). The survey does not contain the data on units bought for most goods and services; therefore we need to use the price data to be able to calculate the inflation rates.

The categories of goods that are surveyed change in line with changing expenditures of households. Oven-ready food was one of the examined categories in 1989, but it disappeared during the 90s as people started to prefer fresh food. Some categories changed quite a lot; communication is one of the examples. In 1990, expenditures on communication were low as it contained only radio and television. Now, there is also post, telephone, mobile phone or internet. Expenses on the whole category of communication are much higher. Generally, the HBS is more detailed nowadays; the number of categories increased over time from 108 in 1990 to almost 300 in last years of the survey.

The HBS is a very good data source for households spending, because business spending is excluded from the survey. It is used during the implementation of social policy of the state, in social and economic research, it is internally used by the CZSO, but also for international comparisons. It has also a number of drawbacks that should be mentioned before proceeding with the analysis.

[3] Quotas are set based on the sample survey of Living Conditions, which is a national modification of European survey: European Union - Statistics on Income and Living Conditions (EU-SILC). This survey is carried out by random selection in order to obtain representative data on the level and structure of income and basic socio-demographic characteristics of households.

Drawbacks of the HBS data

Firstly, Eurostat (2004) adverted to the fact that the HBS is restricted to the population residing in private households. Collective or institutional households (old persons' homes, hospitals, hostels or prisons) are excluded from the survey. They also mention that the population excluded in this way amounts to no more than 2% of the total population, but the effect is more significant for particular groups such as old persons, and certainly the homeless people. It is important to have this in mind when evaluating the overall results for pensioners.

Secondly, Hait & Janský (2014) pointed out that the selected sample might be too small - the ratio of surveyed households to total households was the second lowest among 13 candidate countries in 2004. This issue is very reasonable, but as the sample is not selected randomly, but by intentional quota sampling, we might believe that it reflects well the structure of the Czech population.

Thirdly, Eurostat (2004) pointed out that for some countries there exists a problem of non-response, response rates for 2004 candidate countries ranges from 38% in Malta to 94% in Cyprus.

Lastly, Crawford & Smith (2002) discussed the potential big problem of under- or over-reporting of selected goods and services. The reason behind incorrect reporting can be forgetting, concealment or guilt (the case of alcohol and tobacco). We could compensate for the underestimation in case of alcohol and tobacco, but we are unable to identify other categories that are over or underestimated. We assume in line with other studies that used this dataset that the effect of over or underestimation is not very big, but we have in mind that results for selected categories might be a bit disturbed by this fact.

The treatment of these drawbacks is beyond the scope of this work. The HBS was used in many research papers and the authors try to construct the survey in such way, that it truly reflects the structure of the population. For that reason, we use the data without bigger adjustments.

3.3 Matching Price and Expenditure data

Expenditure categories contained in the HBS data and price indices do not correspond exactly, so consolidation of the data was needed to provide a proper match of the two datasets. We have constructed a conversion table that links variables from HBS data to one of the COICOP categories of the price

data. Generally, the Household Budget Survey has more categories than price data. For example, there are six separate categories for meat products in the HBS data (beef, pork, chicken, other meat, meat in cans and sausages), while only one price category exists in the COICOP (0112 Meat). To link the two, we combine the corresponding expenditure data to form a single item. This procedure is used for many other categories, but it is straightforward as the categories either have similar name or we are able to match the items based on their names.

Unfortunately, we cannot make use of the whole dataset that we have, because of the need to match the two datasets. We have the HBS data for the period 1990-2012, so the analysis is restricted to this period. Consumer price indices are published monthly, but the HBS data only on an annual basis, so we cannot use monthly prices.[4] Another loss comes from the fact that the CPI data are collected from national sources, so there is no use of regional variation of expenditure data in this paper. The CPI is constructed separately for Prague, so comparison of the whole republic and Prague would be the only regional variation possible. We will ignore regional issues; therefore the differences in indices in our results are caused entirely by differences in spending patterns.

The number of variables and the structure of expenditures differ slightly for some years in the sample, so we construct more conversion tables. At the end of matching we obtain the expenditure data where every category of expenditures has one assigned class of the COICOP classification.

[4] Annual averages are used in computations.

Chapter 4

Inflation and its representativeness

4.1 Consumer Price Index (CPI)

Inflation is usually perceived by general public as something very simple, but in reality people do not understand it completely. Inflation is measured by the annual percentage change in the level of prices. The increase differs significantly between goods, so we need to average the rates of the individual goods to obtain an average inflation rate. In other words, inflation measures the change in the cost of living. Unfortunately, calculation of an exact cost of living index is not feasible and we need to use a proxy for it, which is usually a price index. The well known price statistics are Consumer Price Index (CPI), Producer Price Index (PPI) or GDP Deflator. Less common indices are Personal Consumption Expenditures Price Index (PCEPI) used in the US, and the UK also uses Retail Prices Index (RPI).

The official Czech inflation rate is the one published monthly by the Czech Statistical Office (CZSO), the approximation of inflation is Consumer Price Index (CPI) that measures the changing cost of the shopping basket of goods and services purchased by a typical household. CPI can be approximated either by Laspeyres or Paasche index. The official Czech inflation rate is approximated by Laspeyres index, which is defined as:

$$CPI_L = \frac{\sum_{i=1}^{n} P_{i,t} Q_{i,t-1}}{\sum_{i=1}^{n} P_{i,t-1} Q_{i,t-1}} - 1 \tag{4.1}$$

where $P_{i,t}$ is a price of a good i in time t consumed by the whole economy,

$Q_{i,t-1}$ stands for quantity of good i consumed in time $t-1$ and n is the number of goods in the market basket.

The only difference from Paasche index is that Laspeyres uses quantities from the base period $t-1$, while Paasche uses quantities from the period for which the index is computed t. Following equation defines Paasche index:

$$CPI_P = \frac{\sum_{i=1}^{n} P_{i,t} Q_{i,t}}{\sum_{i=1}^{n} P_{i,t-1} Q_{i,t}} - 1 \tag{4.2}$$

where $P_{i,t}$ is a price of a good i in time t consumed by the whole economy, $Q_{i,t}$ stands for quantity of good i consumed in time t and n is the number of goods in the market basket.

The Laspeyres index tends to overstate inflation, the Paasche index tends to understate it. The reason behind is that these two indices do not account for the fact that consumers can react to price changes by changing the quantities of goods and services that they buy. The problem of the overstatement of the Laspeyres index is further described in Subsection 4.2.1. For the purposes of this work, we will consider only CPI as a measure of inflation taking into account all of its drawbacks that are discussed in the Section 4.2.

Quantities needed for calculation of Equation 4.1 are usually not available, however it is possible to obtain the data on expenditures on goods and services. Therefore the Equation 4.1 can be rewritten as:

$$CPI = \frac{\sum_{i=1}^{n} P_{i,t} Q_{i,t} \frac{P_{i,t}}{P_{i,b}}}{\sum_{i=1}^{n} P_{i,t} Q_{i,t} \frac{P_{i,t-1}}{P_{i,b}}} - 1 = \frac{\sum_{i=1}^{n} E_{i,b} \frac{P_{i,t}}{P_{i,b}}}{\sum_{i=1}^{n} E_{i,b} \frac{P_{i,t-1}}{P_{i,b}}} - 1 \tag{4.3}$$

where $E_{i,t} = P_{i,t} Q_{i,t}$ is the aggregate expenditure of the whole economy on good i at time t, $b = t$ is the base period.

4.2 Drawbacks of Laspeyres index

There are two main issues associated with Laspeyres index as a measure of inflation: overstatement of the inflation due to biases (substitution bias, new

items bias, quality bias and outlet bias)[1] and representativeness of the inflation for different types of households. All the biases are shortly described in Subsection 4.2.1, even though they are not the subject of the thesis. The subsection mainly describes how this paper overcomes the problem of substitution bias. Subsection 4.2.2 is devoted to the problem of a typical household which leads us to the average inflation rate representativeness, which is further analysed in Section 4.4.

4.2.1 Fixed-base price index bias

Consumer Price Index is not a measure of a change of true cost of living. It does not take into account switch towards different goods and services, the launch of new goods, the quality change of base goods or the switch towards cheaper outlets. Crawford & Smith (2002) discuss the difference between the cost-of-living price index and fixed-base price index in very comprehensive way. Fixed-base CPI considers purchase of the same combination of goods as in the base period, which is one way to achieve the same level of welfare as in the base period. The second and more effective way to preserve the same level of welfare is to allow for substitution of goods purchased in consequent periods. DeMilner (1981) summarizes the drawbacks of fixed-base price indices and explains that changes in the consumption might result from changing tastes, different levels of income, but most often, they arise from substitution of goods in response to relative price changes. He mentions an example of gasoline price: if the price of gas increases, people reduce their purchases of gas and might substitute usage of cars by the means of public transport.

We know from microeconomic theory that the consumer is rational and maximizes utility; he chooses the combination of goods that minimizes the costs to reach the reference level of utility. Consumer rationality implies that the true cost-of-living index must be smaller or equal to the Laspeyres index. Laspeyres index is therefore the upper bound of the true cost of living index. The difference between the two is the CPI bias. Several studies have tried to assess its value, but their results differ quite significantly. The Table 4.1 adapted from Neves & Sarmento (1997) illustrates the estimated total CPI bias from several studies conducted in the US, Canada and UK.[2]

[1] See Nevest & Sarmento (1997) for detailed analysis of the substitution bias.

[2] Studies from the table are not mentioned in the references, as they were not used for this paper, consult Neves & Sarmento (1997) for the references.

Table 4.1: CPI bias

Study	Year	Country	Bias
Lebow, Robets & Stockton	1992	USA	1.0
Crawford	1993	Canada	0–0.6
Cunningham	1996	UK	0.4–0.8
Boskin	1996	USA	1.1
Lebov & Rudd	2001	USA	0.6

Source: Neves & Sarmento (1997), p. 26.

The value of the CPI biases ranges from 0 to 1.1 percentage point, but all the studies find it significant. Especially Boskin (1996)[3] raised a great wave of interest, because inflation is used to index pensions and social benefits and his result explained to the public that the federal budget increased more than necessary. Disregarding the truth of the number, the Boskin Report has raised a large discussion and drawn the interest to the topic of CPI bias.[4]

Hobijn & Lagakos (2003) points out that substitution bias can be minimized by frequent changes of the base period. The CZSO used to update the base period approximately every five years, but in the last years the base period is updated more often, every two years. If the base period is updated infrequently, CPI does not reflect substitution of goods and services that become relatively expensive for cheaper goods. The US CPI bias was generally considered to be around one percentage point. For that reason we might expect the CPI bias for the Czech Republic to be lower than one percentage point as the base period is being changed two or three times more often than historically in the US. Maybe in reaction to the Boskin Report or in order to decrease the CPI bias, the US practice has changed in 2002 and they started to update the expenditure reference period every two years.[5] Cage et al. (2003) state that the Chained Consumer Price Index was designed to be a closer approximation of the cost-of-living index than previous measures. It would be interesting to analyse whether the CPI bias has changed after the implementation of this schedule and if it would be useful to change the base period more often also for

[3] Boskin Report is formally called "Toward a more accurate measure of the cost of living".

[4] Gordon (2000) and (2006) was the main opponent of the Report and Bureau of Labour Statistics that constructs CPI also reacted in 2006.

[5] BLS (1997) Handbook of Methods, Chapter 17 - Consumer Price Index, p. 33.

the Czech Republic. We do not examine CPI bias in this work, but it would be an interesting topic for further research.

Similarly to Hobijn & Lagakos (2003) and Hait & Janský (2014) this paper eliminates the substitution bias completely, because it updates the base period in each period. Setting $b = t - 1$ to the Equation 4.9 we obtain:

$$\pi_t = \frac{\sum_{i=1}^{n} E_{i,t-1} \frac{P_{i,t}}{P_{i,t-1}}}{\sum_{i=1}^{n} E_{i,t-1} \frac{P_{i,t-1}}{P_{i,t-1}}} - 1 = \frac{\sum_{i=1}^{n} E_{i,t-1} \frac{P_{i,t}}{P_{i,t-1}}}{\sum_{i=1}^{n} E_{i,t-1}} - 1 \tag{4.4}$$

Using Equation 4.4 we do not only reduce the substitution bias, but also the new item bias and the results do not depend on the choice of the base period.

4.2.2 Typical household

The second issue is the word "typical household" in the CPI definition. Naturally, there is no such thing as typical household. Common sense tells us that spending patterns of young families differ notably from those of a family of two pensioners. Leicester & al. (2008) describes an even better example of household differences: as other groups of goods, alcohol and tobacco have assigned weights in the consumer price index and those weights will be too high for households that do not smoke or drink, but on the other hand they are too low for households that smoke and drink a lot. Households without a car do not spend anything on fuel; households without pets do not spend money on pet food and so on.

For the purpose of the thesis, we define also household specific inflation rate similarly to Equation 4.4:

$$\pi_{h,t} = \frac{\sum_{i=1}^{n} E_{h,i,t-1} \frac{P_{i,t}}{P_{i,t-1}}}{\sum_{i=1}^{n} E_{h,i,t-1}} - 1 \tag{4.5}$$

where $E_{h,i,t}$ is the expenditure of household h for the good i in time t.

As Hobijn & Lagakos (2003) discuss, we are interested in household specific expenditures and household specific price changes. Unfortunately, we are not able to observe the specific prices that households pay for groups of items and we assume that all households face the same price increases for each group of

goods. Most of the empirical research makes this assumption. An example of the study that examined whether different groups of households face different price changes is Berndt et al. (1998) who found out that elderly people face similar price changes in prescription drugs as young people, even though they use very different drugs. For this reason we will focus only on household specific expenditures and we will make the assumption of equal price increases for all households.

4.3 Official vs calculated inflation

The average inflation that is used in the thesis differs from the official inflation published by the CZSO. The annual inflation as defined by the CZSO is expressed by the increase in the average annual consumer price index - the percentage change in the average price level for the last twelve months compared with the average of twelve previous months.[6]

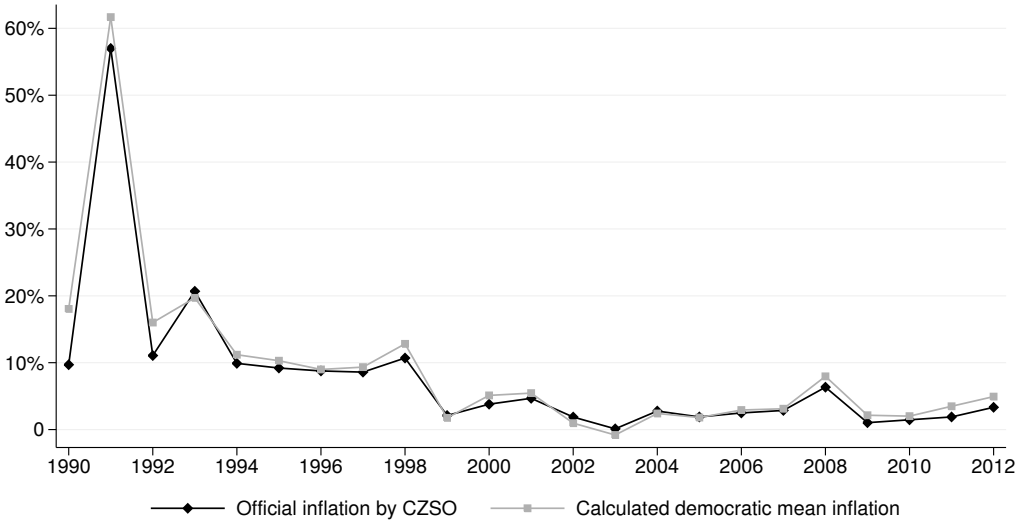
The average inflation used in the paper is based on individual inflation rates of households in our sample. Firstly, we have calculated household price indices based on the expenditure data from the HBS. Then we have calculated the inflation for households that were present in the sample in two consecutive years. The annual inflation was calculated using the average of specific household inflations for particular year and we do not take into account coefficients that the CZSO uses for calculation of the official inflation.

Figure 4.1 plots both inflation rates: the calculated and the official one. There are significant differences between two inflation rates during the whole period. For that reason we do not compare our results for subgroup inflation rates with the official inflation from the CZSO, but with the inflation of other groups in the sample or with our average inflation. Our calculated inflation is on average higher than the official inflation published by the CZSO. The average calculated inflation rate for the period of 1990–2012 is 9.19% while the official average inflation rate in this period was 7.93%.

Inflation is perceived mostly as a macroeconomic variable that affects all groups of people in the same way. If this underlying assumption is not true, policy makers should consider constructing more types of consumer price indices, which would differ for regions or for groups of households. The aim of the

[6] For detailed description of construction of price indices, selection of goods and adjustments, please refer to methodical guidebook for users of CPI, published by CZSO in 2015.

Figure 4.1: Offcial vs calculated inflation



Source: CZSO, HBS data and author's calculations

Section 4.4 is to find out whether the differences between individual households are significant enough to draw the attention of policy makers, in other words how representative the official inflation is.

4.4 Average inflation rate representativeness

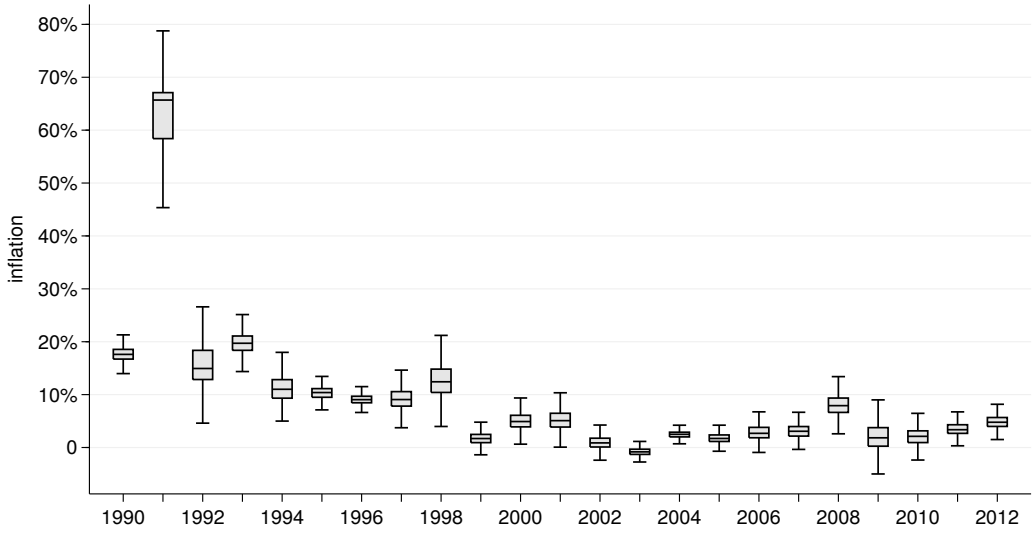
To determine the representativeness of the average inflation rate it is important to look at the variation of inflation across households. No variation would mean that the mean inflation is a perfect measure of inflation for every household. On the other hand, the case of very high variation would imply that almost no households have the mean inflation rate. These are two extreme cases, but the reality is as always somewhere in between. The variation is caused by different inflation rates of particular goods that are represented in the consumer basket of each household in a different proportion.

This subsection will examine the representativeness of the average inflation on a sample of Czech households for 1990–2012. The data that we use for the analysis are described in detail in Chapter 3.

Figure 4.2 indicates the degree of the variation of inflation. We see that it is relatively high and it is reasonable to examine it in detail. The line in the box depicts the mean inflation; the whole box covers 25th–75th interval and lines the whole distribution. Outliers that were depicted on the graph are

omitted for better orientation. As expected, variation is higher in times of high inflation and in years when the economy was not stabilized (post-communist period or years after the financial crisis 2008–2009). Generally, the variation is not skewed, it is only slightly skewed in 1991 and 1992, in years after price liberalisation. Interesting examples are the years with mean inflation below 5% when some households experienced deflation. An extreme example is year 2009 when the average inflation rate was 1.88%. The variation was very high; inflation ranged from -7% to 9%. Without seeing the dispersion, we would say that it was a year with low inflation, but can we say the same thing for the whole population after seeing the distribution?

Figure 4.2: Distribution of the annual inflation



Source: CZSO, HBS data and author's calculations

Figure 4.2 showed that there is a relatively wide variation in the dispersion of inflation. We may ask how representative the average inflation is and to measure the representativeness we ask if the inflation of a particular household is close to the average rate. Leicester et al. (2008) proposed two alternative summary measures of the "closeness".[7] The first measure is defined as the answer to the question how many households are less than 25% away from the mean inflation rate. Formally defined as:

[7] Same measures were used by Hait & Janský (2014), Crawford & Smith (2002) and Artsev et al. (2006).

$$\overline{\pi} \geq 0 : \pi_h \in (0.75\overline{\pi}, 1.25\overline{\pi}) \Leftrightarrow \pi_h \in C \tag{4.6}$$

$$\overline{\pi} < 0 : \pi_h \in (1.25\overline{\pi}, 0.75\overline{\pi}) \Leftrightarrow \pi_h \in C \tag{4.7}$$

where C is the subset of households that are "close" to the average inflation, π_h is a specific inflation rate for households h and $\overline{\pi}$ is the mean inflation rate.

The second measure asks how many households have the inflation rate less than one percentage point away from the mean inflation. Formally:

$$\pi_h \in (\overline{\pi} - 1\%, \overline{\pi} + 1\%) \Leftrightarrow \pi_h \in C \tag{4.8}$$

Each measure will yield different results depending on the distribution of inflation in a particular year. For the first measure many households are close to the mean when the inflation is high.[8] The opposite holds for the second measure; households are close to the mean in cases of low inflation.[9]

Figure 4.3 plots the first measure of closeness (black line) together with the mean inflation rate (grey line).[10] Leicester et al. (2008) depict on graph households that are far from the mean, but we instead plot a percentage of households that are close to the inflation for both measures as we are interested in the representativeness of inflation rate. As expected, representativeness is higher for years with high inflation. It remained over 50% for the whole period from 1990 to 1998. There are two sharp falls in the after-2000 period: the first is in 2002 when inflation was below 1% and it persisted until 2003 when the Czech Republic experienced deflation (representativeness is very low for those two years; it is only 15.73% and 22.56%). A similar decrease was in 2009 when only 16.09% of households was close to the mean inflation rate of 2.14% based on Measure 1. The positive relation between the average inflation rate and households that are within 25% around the mean is clearly visible, but it is not perfect. Our results for Measure 1 are in line with the findings of Leicester et al. (2008) for the UK, who have found the same pattern. The average representativeness across the whole period is only 58.2%. Overall, we can say

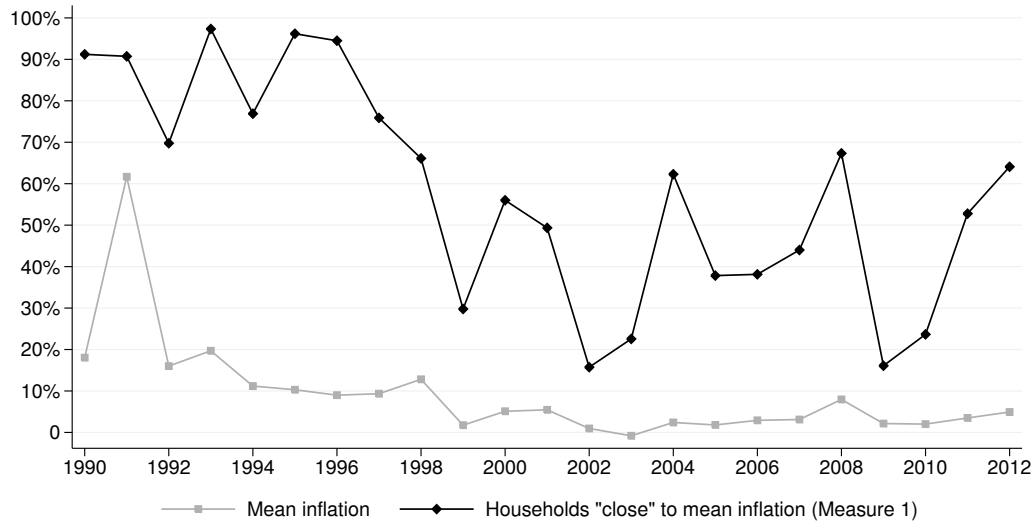
[8] For the inflation rate of 1% it will count only households within the range (0.75%,1.25%), while for the inflation rate of 10% the band will be broader: (7.5%,12.5%).

[9] If we have inflation rate of 1% the range is (0%,2%), while for the inflation rate of 10% it is (9%,11%).

[10] Figures for all years are summarized in a table in Appendix B.

that the representativeness for Measure 1 is quite low and suggests that the use of group specific inflation rates might be reasonable.

Figure 4.3: Measure 1: Percentage of households that are 25% around the mean inflation rate



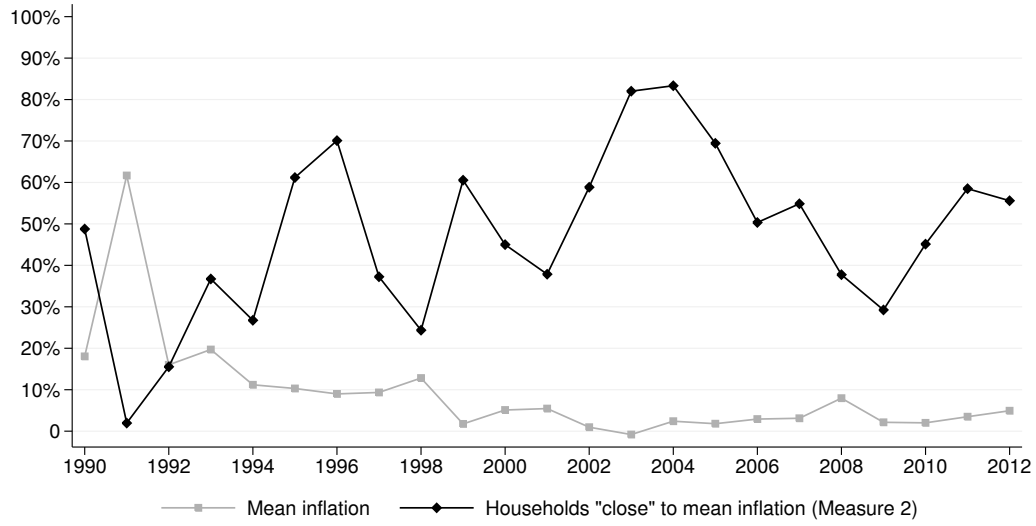
Source: CZSO, HBS data and author's calculations

The second measure is not only used by Leicester et al. (2008), but also by Crawford & Smith (2002) and Artsev et al. (2006). Figure 4.4 shows the second measure, the percentage of households with the inflation rate more than one percentage point from the mean.[11] For this measure, we expect a negative relationship; representativeness should be low for years with higher inflation. Our results are in line with expectations, but there are again some exceptions. The lowest representativeness is 1.96% in 1991, when the inflation was extremely high after the liberalisation of prices. On the other hand, the highest representativeness of the average inflation rate appeared in 2003 when there was deflation and it remained that high also in 2004. The average representativeness over the whole period is even lower than in the case of Measure 1; it is only 47.4%.

There is one interesting and unexpected finding from Figure 4.3 and Figure 4.4. It seems that when there is a sharp increase or decrease in the representativeness of the average inflation, it persists also the following year (if the mean inflation is not too different). It seems to be specific to Czech results.

[11] Figures for all years are summarized in a table in Appendix B.

Figure 4.4: Measure 2: Percentage of households one percentage point around the mean inflation rate



Source: CZSO, HBS data and author's calculations

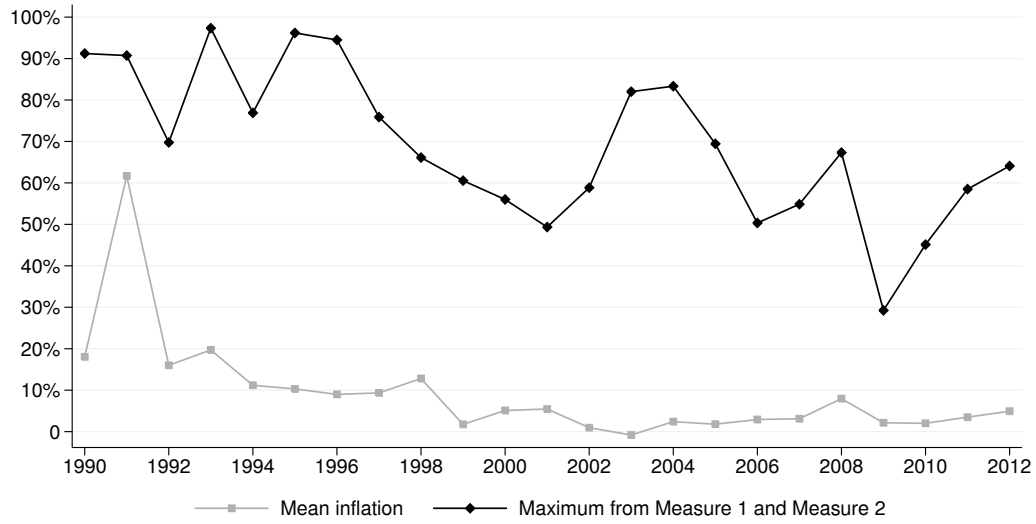
Findings from the UK do not show this persistence; representativeness there reflects the changes in inflation better.

We have seen that both Measure 1 and Measure 2 can be used as measures of "closeness" to the average inflation, but Measure 1 shows high closeness if the inflation is high (the range is bigger) and Measure 2 if the inflation is lower (+/- one percentage point). Therefore we take maximum from these two values as the most relevant measure of closeness. We see on Figure 4.5 that the closeness to mean inflation is higher and it falls below 50% only in 2009 and 2010. Many households are close to the average inflation rate in times of inflation rate higher than 10%, on the other hand, more households are far from the mean if the inflation rate is close to zero. The average representativeness over the whole period is 69.1%, which is more reasonable than taking Measure 1 or Measure 2 separately.

The last measure of the average inflation rate representativeness was also presented by Leicester et al. (2008). This concept is the mean absolute deviation (MAD), which calculates the average difference between the inflation rate of each household and the mean inflation rate. The following is the formal definition:

$$MAD = \frac{\sum_h |\pi_h - \overline{\pi}|}{N} \tag{4.9}$$

Figure 4.5: Maximum from Measure 1 and Measure 2



Source: CZSO, HBS data and author's calculations

where π_h is a specific inflation rate for households h and $\overline{\pi}$ is mean inflation rate and N is number of households in the sample for a particular year.

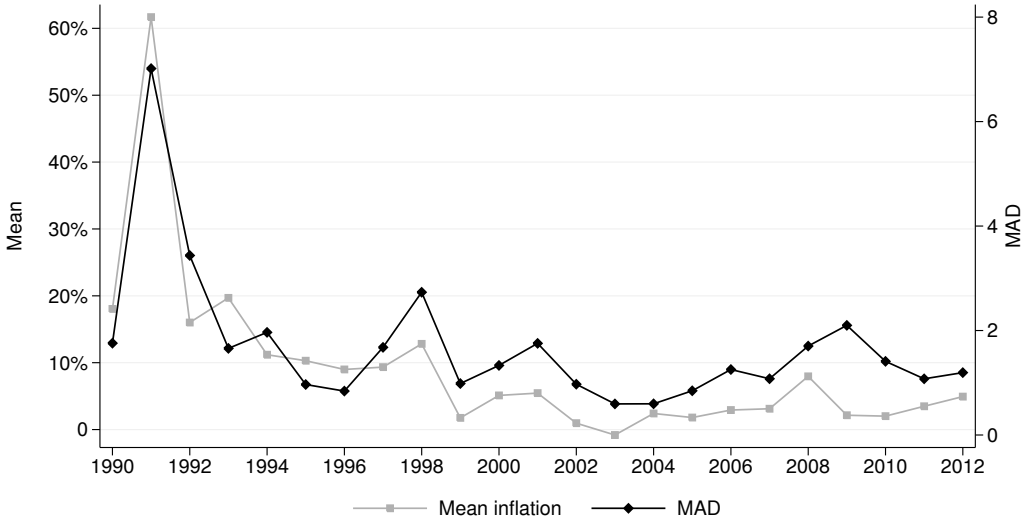
The advantage of mean absolute deviation in comparison to Measure 1 and Measure 2 is that we do not need to set any threshold (25 per cent for measure or one percentage point away from the mean), because MAD summarizes the whole distribution instead of comparing households with given threshold.

Before proceeding to Figure 4.6 it is useful to understand the concept of mean absolute deviation. If MAD is two, it means that households are on average two percentage points from the mean inflation rate. [12]

In Figure 4.6, the average inflation rate is depicted on the left-hand axis and MAD on the right-hand axis. We would expect there to be a positive relationship between the two, but results of MAD follow the changes in inflation almost perfectly. Surprisingly, the correlation is also high for years with extremely high inflation or for a year with deflation. Concretely, in 1991 when the average inflation was 61.68% MAD reached 7.02, which means that households were on average seven percentage points from the mean. In 2005 the inflation was 2.41% and MAD was 0.66, households were on average 0.66 of a percentage point from the mean. This measure is probably the most illustrative example of the non-representativeness of the average inflation rate for all households.

[12] Figures for all years are summarized in a table in Appendix B.

Figure 4.6: Mean absolute deviation (MAD)



Source: CZSO, HBS data and author's calculations

Based on the results from this section, we see that inflation does not vary only over time, but also across households. The question is what the causes of variation are and if we can find some groups that would have similar inflation significantly different from the average. In Chapter 5 we will construct specific subgroup inflation rates for selected groups that are expected to have similar spending patterns.

Chapter 5

Subgroup inflation rates

We have seen in Chapter 4 that price indices might be biased. Analysis of the average inflation rate representativeness has shown that percentage of households close to the official inflation rate can be very low for some years. This chapter will examine whether the inflation differences between households are random or whether similar groups (more precisely groups that are expected to have similar spending patterns) systematically have inflation rate different from the mean inflation rate.

The first section describes the differences between two possible approaches to the aggregation of household inflation to construct subgroup inflation rates, Section 5.2 compares inflation rates for income deciles in our sample and Section 5.3 compares inflation rates of pensioners and non-pensioners. In Section 5.4 we consider the number of people in a household, in Section 5.5 the number of children and in Section 5.6 we construct subgroup inflation rates for households from a city or from a village.

5.1 Plutocratic vs democratic mean

Before proceeding to the construction of subgroup inflation rates, we need to consider how to aggregate individual household inflation rates.[1] Two broadly used approaches how to calculate mean inflation rates exist: democratic and plutocratic approach. The difference between the two is discussed in detail in Crawford & Smith (2002) and Artsev et al. (2006). Hait & Janský (2014) also

[1] Household specific inflation rates were defined in Subsection 4.2.2.

compute the average inflation rates for the Czech Republic using both methods and compare them.

Democratic average weights households equally; all households have the same importance. This unweighted arithmetic mean inflation is formally defined as:

$$D_{x,i} = \frac{\sum_{i=1}^{m} \pi_{h,i}}{m} \tag{5.1}$$

where $\pi_{h,i}$ is the specific inflation of household h and m is number of all households in the subgroup x.

Plutocratic average assigns weights to households according to their total expenditures; therefore richer households receive more weight. This weighted arithmetic mean is formally defined as:

$$P_{x,i} = \frac{\sum_{i=1}^{n} E_{x,i,t-1} \frac{P_{i,t}}{P_{i,t-1}}}{\sum_{i=1}^{n} E_{x,i,t-1}} - 1 \tag{5.2}$$

where $E_{x,i,t}$ is expenditure of group x for the good i in time t and $P_{i,t}$ is a price of a good i in time t consumed by the whole population.

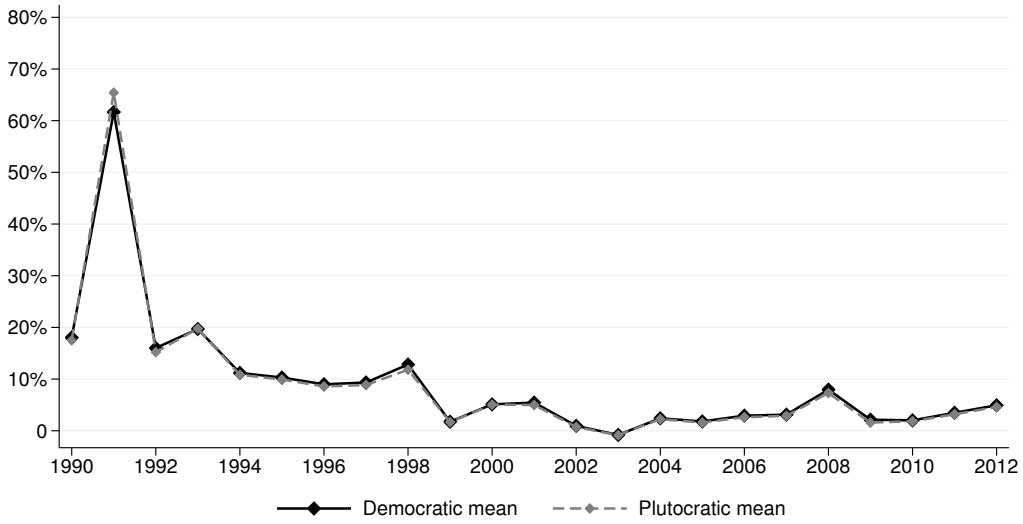
Both types of aggregation are equally valid from the economic point of view.[2] Artsev et al. (2006) mention the main advantage of plutocratic average: it does not require computation of specific households' inflation rates, it uses total expenditure of the whole group for the particular good. As computation of specific household rates is part of this work, we prefer to use democratic average inflation rate, which gives the same weight to every household. Plutocratic weights are used by almost all countries to construct consumer price indices. Ooisthuizen (2007) points out that top quintile in South Africa accounts for more than 70% of the CPI. He also explains that the use of plutocratic weights is reasonable for construction of CPI as it is used as the indicator of the whole economy and it is therefore preferable to weigh the index according to total consumer expenditures. On the other hand, democratic average is more suitable

[2] Prais (1959) is well known for being one of the first to discuss difference between weighted and unweighted aggregate inflation rates.

for this work as it involves equally middle- or low-income households and we are interested in constructing inflation rates for specific groups.

Figure 5.1 plots both aggregate measures of the average inflation rate for the total sample. It seems from the graph, that both measures are very similar, but we need to consider any small differences (0.5 of a percentage point might be a significant difference in case of inflation). It is hard to see the differences with the wide range on y axis (which is caused by a peak in 1991), therefore the difference is plotted on a separate graph.

Figure 5.1: Plutocratic vs democratic mean



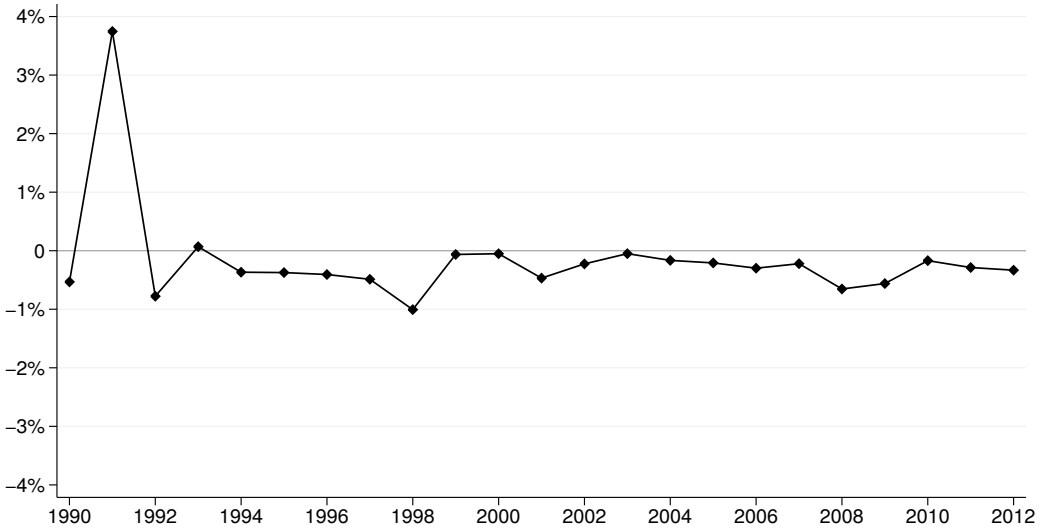
Source: CZSO, HBS data and author's calculations

Difference between plutocratic and democratic mean is usually referred to as plutocratic bias. The idea behind is following: plutocratic mean gives more weight to richer households, so if rich households experience higher inflation than low-income households, plutocratic bias is above zero. Plutocratic bias lower than zero, on the other hand, indicates that low-income households experienced higher inflation than high-income households in that year. Plutocratic bias also indicates whether the rich and poor experience significantly different inflation rates. If there are systematically large differences in inflation rates between income groups, plutocratic bias tends to be larger than if all households experience very similar inflation rates.

Figure 5.2 plots the difference between plutocratic and democratic mean values. We see that the plutocratic bias is negative for all years except for 1991 and 1993. These two years were specific years during the Czech transition

period when the overall inflation was very high. We have already described exceptional events in these years at the end of Section 3.1, where we describe the CPI data.[3] Plutocratic bias is negative for all other years in the period of 1990–2012 and it is in between -1 percentage point and zero. It confirms that low-income groups experience systematically a higher inflation rate over the period.

Figure 5.2: Plutocratic bias



Source: CZSO, HBS data and author's calculations

Negative plutocratic bias in Figure 5.2 together with low inflation rate representativeness analysed in Section 4.4 justify the construction of specific inflation rates for groups with similar spending patterns. We will construct subgroup inflation rates for low- and high-income households, households with many children or no children, pensioners, households with one person against many and households living in a city or in a village in the following subsections. We use democratic average in this chapter as we want to assign the same weight to every household. Figures for all subgroup inflation rates are summarized in **??**.

5.2 Inflation rates by income

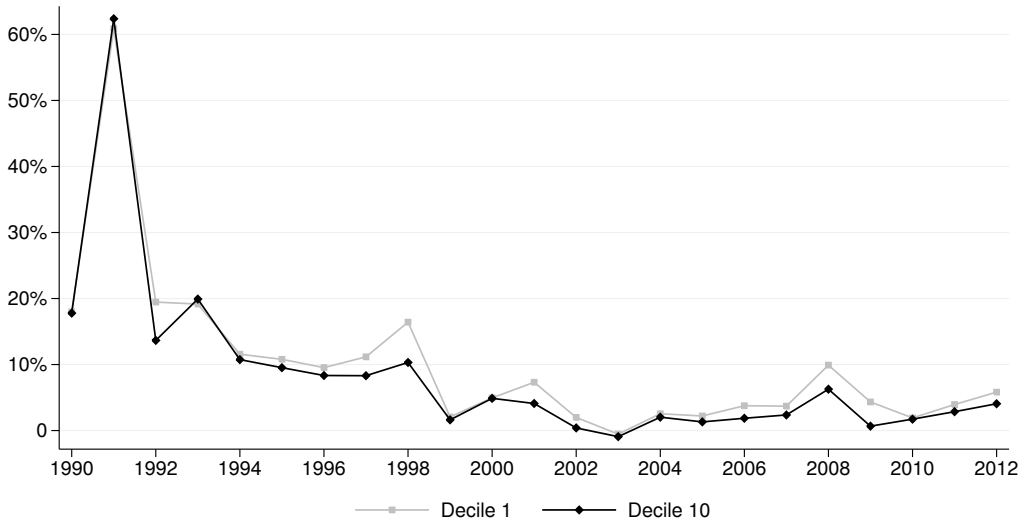
In this subsection, we can relate the direction of the plutocratic bias from Section 5.1 to the inflation of low- and high-income households. In years when

[3] Price liberalisation in 1991 and introduction of value added tax in 1993.

plutocratic bias is very close to zero, there should be no difference between inflation of rich and poor. When the plutocratic bias is positive, high-income households should face higher inflation rates than low-income households. On the other hand, if the plutocratic bias is negative, low-income households face higher inflation than the rich part of the population.

Figure 5.3 shows the inflation rates for decile 1 (poorest 10% of households) and for decile 10 (richest 10% of households) for the whole period since 1990 to 2012. Inflation rates for rich and poor are significantly different from each other for most years and low-income households face consistently higher inflation than rich households. There are two exceptions in the beginning of the 90s: in 1991 and 1993, the richest decile had higher inflation, however the differences between the two were small relative to differences since 1995. Looking back to Figure 5.2 with plutocratic bias, we see that its value is positive only for above mentioned years 1991 and 1993. Since the artificially high inflation in 1991 forces us to have to wide y axis, we will always show subgroup inflation rates on two graphs, one for the whole period from 1990–2012, the other one only for the period since 1999, when the inflation rate is 10% at maximum.

Figure 5.3: Inflation rates by income decile 1990–2012

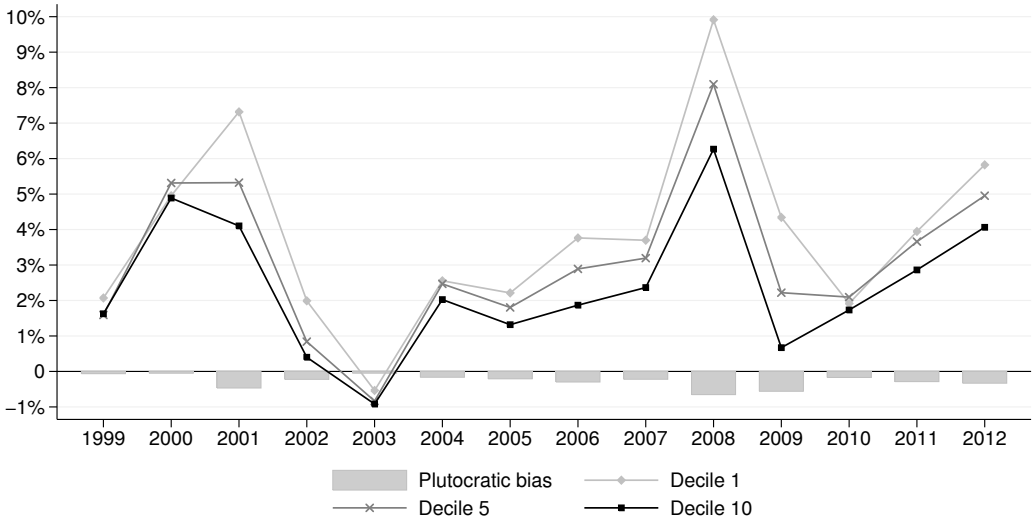


Source: CZSO, HBS data and author's calculations

Figure 5.4 shows the inflation rates for 1999–2012 together with plutocratic bias to see the relation of the bias to the difference of decile 1 and decile 10. We can see that the size of plutocratic bias is very strongly related to the difference between inflation rates of decile 1 and decile 10. Plutocratic bias

is high in three years, one in 2001, the second in 2008 and the third in 2009. The inflation was relatively high in these years and the differences between low- and high-income groups of households were 3.21, 3.65 and 3.67 percentage points respectively. Even though we expected the lowest decile to have higher inflation, it is surprising to obtain such clear results (the ranking of the two deciles often changes in studies conducted in different countries). Decile 5 is in between the values of decile 1 and decile 10 for all years except for 2000 and 2010 when the inflation of decile 5 is the highest. We now compare our results with other studies that constructed subgroup price indices based on income.

Figure 5.4: Inflation rates by income decile 1999–2012 (plutocratic bias)



Source: CZSO, HBS data and author's calculations

Lieu et al. (2004) presented, on empirical evidence from Taiwan, results for years 1993–1996. The lowest income level group had significantly higher inflation for three out of four years, but the differences were smaller than in our case. The period examined was quite short, so we cannot make a conclusion about the sign of the difference.

Leicester et al. (2006) analysed households of pensioners in the UK. As a part of the analysis, he divided them into the poorest pensioners and non-pensioners (households in the poorest 20% of population) and the richest pensioners (households in the richest 20% of the population) and compared their inflation for 2006–2008. The examined period was again really short and the

ranking has changed even during these three years (poor pensioners faced above average inflation, but they had the lowest inflation the next year).

Ooisthuizen (2007) showed on data from 1998 to 2006 that in South Africa decile 10 had relatively balanced inflation, while decile 1 fluctuated around values of the richest decile.

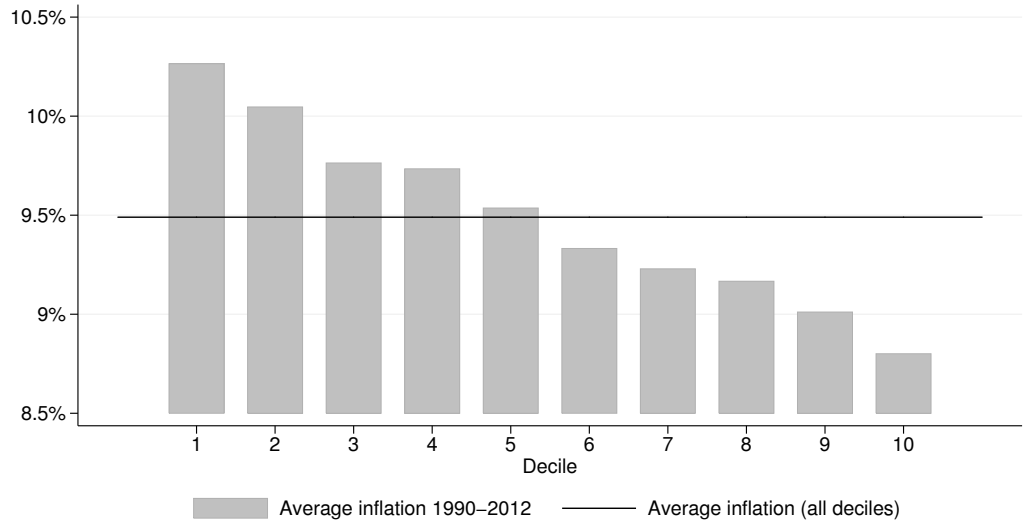
Similar fluctuations are visible on the graph of Crawford & Smith (2002) on the UK data. Unlike others, this paper also calculates and compares annual average inflation rate for all deciles over the whole period. It shows that households with lower income faced on average lower inflation in 1976–2012.

An analysis of the Czech data was done only by Hait & Janský (2014), who compared low-income households with the aggregate population yielding similar results as our analysis. Most importantly, they also found the inflation of low-income households higher than the aggregate average inflation for all years except for 1999 and 2001, when it was the same (1999) or lower (2001). Since we do not compare low-income households with the whole population, but with the rich decile, the results can differ.

Similarly to Crawford & Smith (2002) we also show the annual average inflation rate (geometric mean) for all income deciles over the period in Figure 5.5 and we compare it with the average population inflation rate (9.49%). Not surprisingly, we get completely different results than Crawford & Smith (2002). Five deciles with lower income had more than average inflation rate while five deciles with higher income had lower inflation. Hait & Janský (2014) explain it by the fact that most often poor households are less able to substitute the goods and services they buy if their price increases.

This section has shown that Czech low-income households face systematically higher inflation than households with a high income. The difference between poor and rich ranges from -1.51 to 6.11 percentage points, the average difference between the two is 1.53 percentage points (including two years with negative difference). An example of significant difference in year with low inflation is 2005 when the estimated inflation rate of decile 1 is 2.21%, but the inflation rate of decile 10 is only 1.32%. If we consider that inflation is relatively low in recent years (less than 5%), the average difference of 1.53 percentage points is not negligible and it would be reasonable to construct specific group inflation rates for different income deciles.

Figure 5.5: Average inflation rate over the period 1990–2012



Source: CZSO, HBS data and author's calculations

5.3 Inflation rates of pensioners

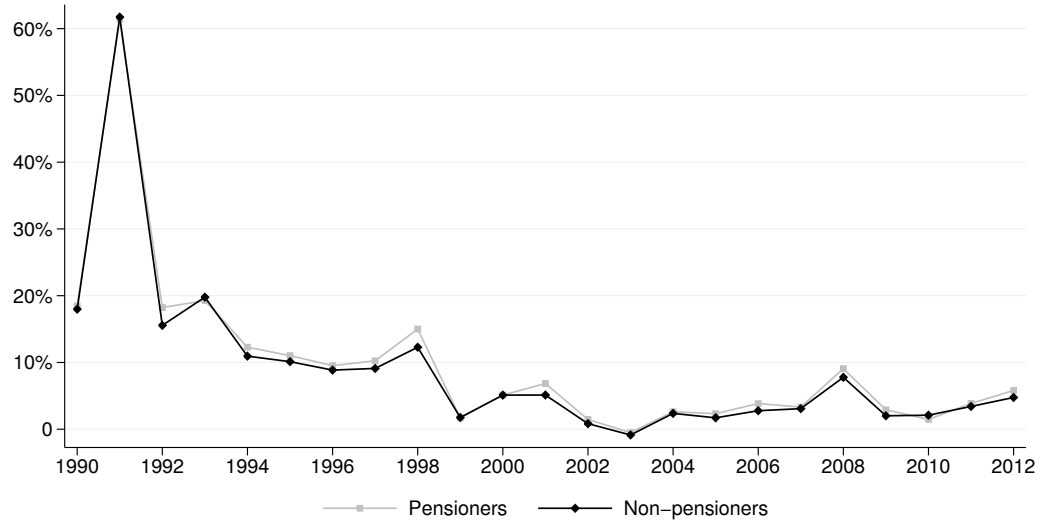
Many studies focus on inflation of pensioners and non-pensioners. Pensioner household is defined for the purpose of this chapter as a family of non-working pensioners receiving a pension.[4] An indicator that there might be a difference between average inflation for the whole population and for pensioners is the fact that the CZSO constructs a specific consumer price index only for two specific groups of people: for pensioners and for people living in Prague.

We can see at first sight in Figure 5.6 that the differences for pensioners and non-pensioners are smaller than they were in the case of groups divided according to income. Nevertheless, the difference exists and pensioners had higher inflation rate than non-pensioners for 19 out of 23 years within the period. The biggest difference is equal to 2.73 percentage points in 1998, the average of differences for the whole period is 0.74 of a percentage point (relatively low number in comparison to 1.59 percentage points for deciles 1 and 10). Figure 5.7 plots the same thing for 1999–2012, but it is more visible that the difference is relatively high.

Results in existing literature are similarly varied as they were in the case of income subgroups. Leicester et al. (2006) showed that pensioners have more

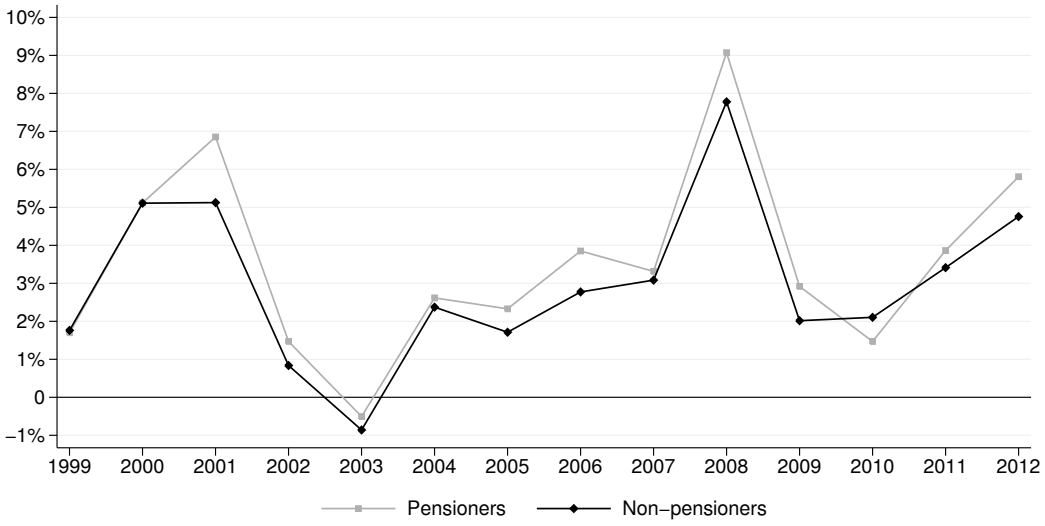
[4] We intentionally exclude households of pensioners raising children or households of pensioners living with any economically active person.

Figure 5.6: Inflation rates for pensioners 1990–2012



Source: CZSO, HBS data and author's calculations

Figure 5.7: Inflation rates for pensioners 1999–2012



Source: CZSO, HBS data and author's calculations

stable inflation in recent years and non-pensioners' inflation is volatile; it is fluctuating around pensioners' inflation rate. They explain it by volatile mortgage interest payments that are part of the consumer basket of non-pensioners only.

In Taiwan, Lieu et al. (2004) showed that pensioners have slightly higher inflation than non-pensioners in 1992–1996, but the difference is negligible in comparison to Czech results.

Hobijn & Lagakos (2003) analysed the US data and found out that in 1987–2001 the inflation for pensioners was usually higher than for the rest of the population. They conclude that it is caused mainly by the rising price of health care.

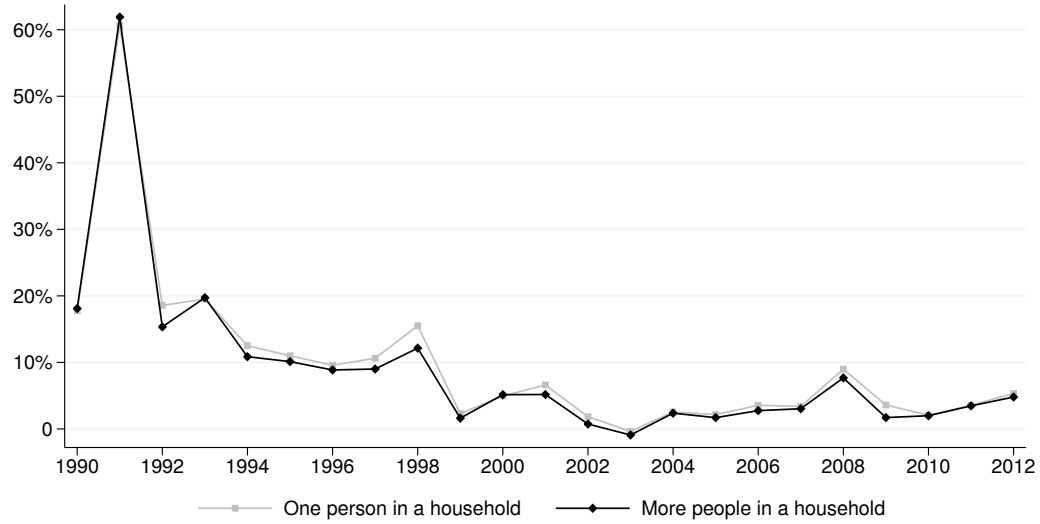
We have shown that Czech pensioners face higher inflation than the rest of the population. The difference is smaller than for low-income households, but it is much higher than what the results from other countries show. Increase of price is an especially important question for pensioners as indexation of pensions is linked to the official average inflation of the whole population. Based on our results, we can say that pensioners' pension would increase more if it was indexed to specific pensioners' inflation rate.

5.4 Inflation rates by the household size

The next subgroup index that we construct divides households into two groups: households of one person and households of more than one person. None of our reference studies have constructed such a subgroup, but Figure 5.8 shows that the differences are of a similar size as differences of pensioners with non-pensioners. We see that person living in a household alone faces higher inflation than bigger households for 19 out of 23 analysed years. The average inflation rate for one-person household for the whole period is 0.82 percentage points higher than for bigger households (it is 0.74 of a percentage point higher for pensioners than for non-pensioners).

Figure 5.9 again restricts the period for 1999–2012. As we expect one-person households to be correlated with pensioners (many households of a single person are pensioners), we plot the values where pensioners are included (solid line) and compare them with values where pensioners' households are excluded (dashed line). It changes the value of households with more people only slightly, but the value of one-person households is significantly lower according to our expectations. But the difference between households of different size is still

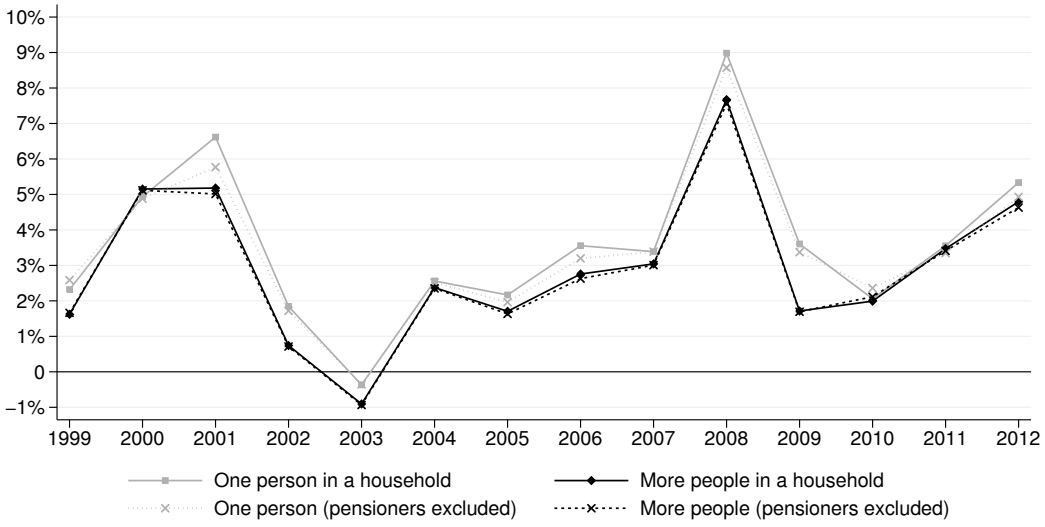
Figure 5.8: Inflation rates by household size 1990–2012



Source: CZSO, HBS data and author's calculations

significant even if pensioners are excluded (the average difference for 1990–2012 is 0.82 of a percentage point if pensioners are included and 0.59 of a percentage point if they are excluded).

Figure 5.9: Inflation rates by household size 1999–2012



Source: CZSO, HBS data and author's calculations

The only study that compared the inflation rates according to household size was Leicester et al. (2006). They analysed only elderly people, but they divided them by family structure to single female, single male, couples and

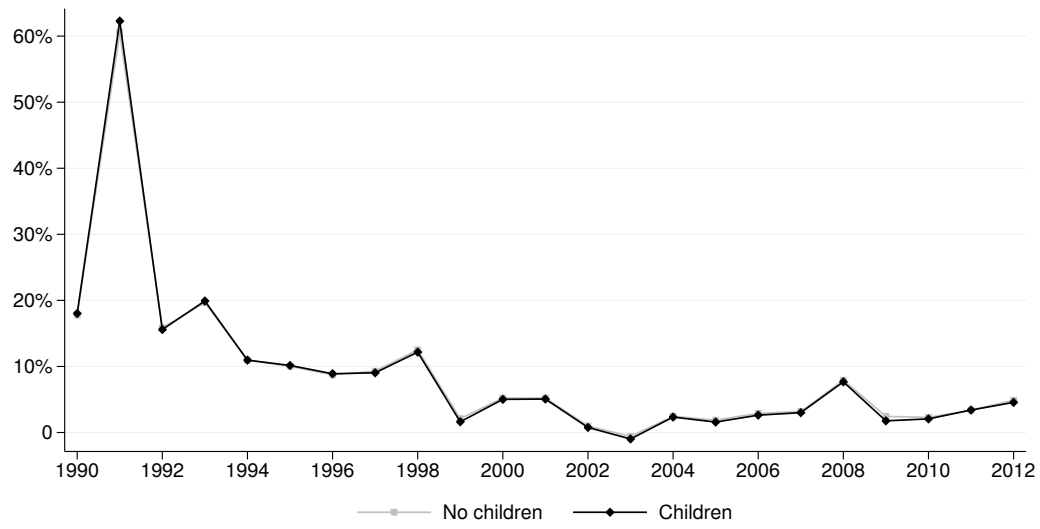
other. They did not find significant differences between males and females and concluded that single pensioners had higher than average inflation.

These results are in line with our results. We conclude that households, where a single person lives, face systematically higher inflation than households of two or more people.

5.5 Inflation rates by the number of children

Figure 5.10 plots the annual average inflation rates for households with a dependent child or children and households without any children. We exclude pensioners' households and construct the subgroup inflation rates only for economically active families. The difference is at first sight much smaller than it was in the case of previous subgroups. Surprisingly families with children do not face higher inflation, the difference between families with and without children is not significant (families without children face on average 0.09 of a percentage point higher annual inflation than families with children).

Figure 5.10: Inflation rates by the number of children 1990–2012

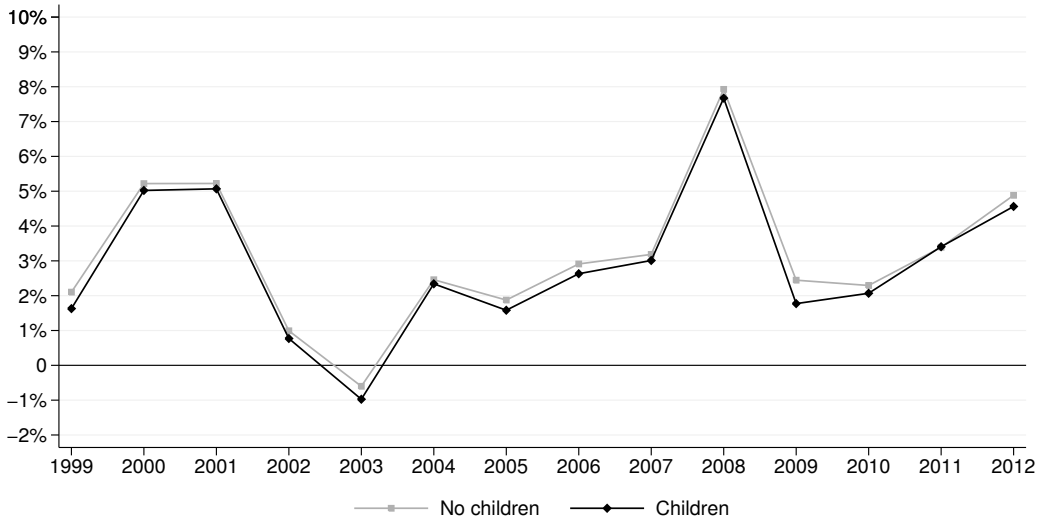


Source: CZSO, HBS data and author's calculations

Figure 5.11 shows that the difference is really small, but it becomes significant for some years in the period of 1999–2012. Families without children face consistently higher inflation except for 2011. Even though the annual average difference for the shorter period is 0.27 of a percentage point, we cannot make any conclusions from it, since we cannot exclude the option that it is caused

by other factors than different spending patterns of families with and without children.

Figure 5.11: Inflation rates by the number of children 1999–2012



Source: CZSO, HBS data and author's calculations

Crawford & Smith (2002) came to the same conclusion about the UK data; there was no significant difference between the average inflation rates experienced by households with and without children in 1976–2000.

Idson & Miller (1999) found that US families with children experienced relatively lower inflation rates than families without children during the 1968–1987 period. They explained it by the fact that families with children are usually younger, and younger families have lower rates of inflation.

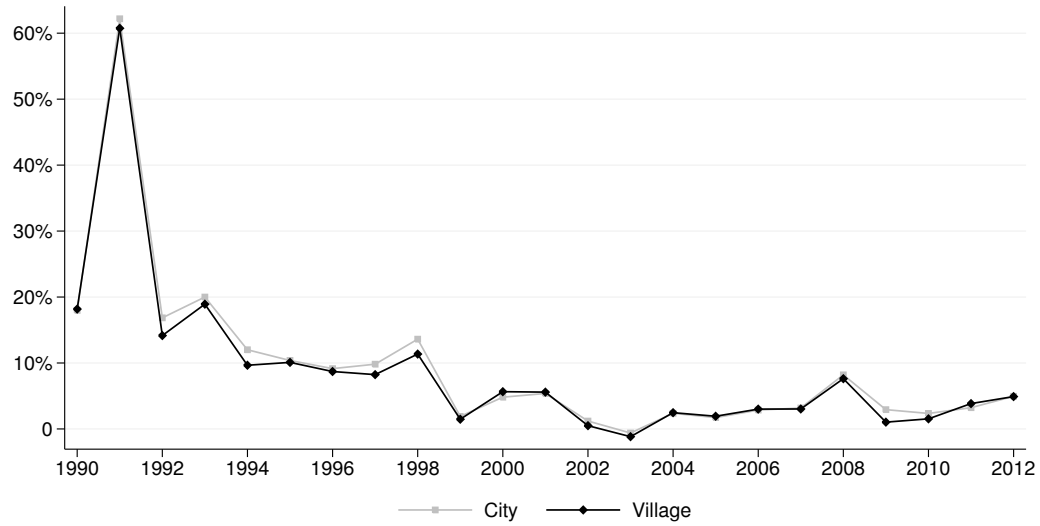
Artsev et al. (2006) showed that the difference in inflation rates for households with and without children is significant for Israel, but its ranking changed frequently and examined period was relatively short (1999–2005).

5.6 Inflation rates by the size of commune

The last subgroup inflation rates are constructed for households living in a city (more than 10,000 inhabitants) or in town or small village (less than 10,000 inhabitants). Figure 5.12 plots both groups on a graph and we see that the urban population faces higher average inflation than households living in rural areas or smaller towns. The difference is significant for most years, but the ranking changes relatively often (households in a city faced higher average

inflation sixteen times and households in a rural areas seven times). Inflation of households in urban areas within the period of 1990–2012 is higher on average by 0.65 of a percentage point every year than inflation of rural area households.

Figure 5.12: Inflation rates by the size of commune 1990–2012



Source: CZSO, HBS data and author's calculations

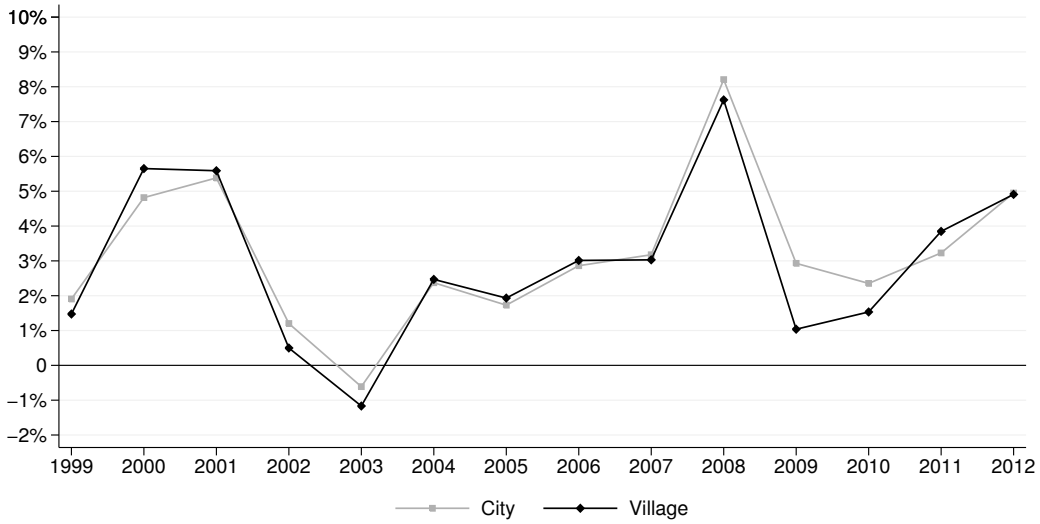
Figure 5.13 plots shorter period from 1999 to 2012. A ranking of the two average inflations is changing throughout the period; the average difference is only 0.22 of a percentage point (very low in comparison to 0.65 of a percentage point for the whole period 1990–2012). The difference was more significant and clear in the 90s; we cannot make any conclusions for the modern period since 1999 as the ranking is volatile.

Murphy & Garvey (2004) analysed urban and rural price indices for the lowest income groups for 1989–2001 on Irish data. They found that all the indices were similar until 1996, but since then prices grew more for the urban poor group.

Artsev et al. (2006) on the other hand, showed on Israeli data that the average inflation for rural areas was slightly higher than for urban areas in 1999–2005.

This chapter showed that the differences in inflation rates between groups of households with specific characteristics exist, the sign of the differences usually does not change over time and the differences are not negligible. Comparison of our results with existing studies on the topic from other countries showed that the difference in inflation rates for groups of households are higher in the Czech

Figure 5.13: Inflation rates by the size of commune 1999–2012



Source: CZSO, HBS data and author's calculations

Republic and what is even more important - the sign of the difference does not change over time. Higher inflation rate applies particularly to vulnerable groups (pensioners or households with lower income), policy makers should consider it and it might be reasonable to construct subgroup price indices and use them for different purposes.

Chapter 6

Empirical analysis

Every household in our sample has specific characteristics such as the level of income, number of children etc. In Chapter 5, we constructed subgroup inflation rates for households that share one common characteristic. Regression analysis allows to explicitly control for many factors that simultaneously affect the dependent variable. This chapter enables us to explain the effect of a household characteristic (explanatory variable) on the inflation (explained variable) holding all other variables that might affect the explained variable constant (ceteris paribus effect). We can compare the results of these two chapters, but we must have in mind this difference. Through this chapter, we refer to Chapter 5 as "qualitative analysis" and Chapter 6 as "quantitative analysis".

By the end of the chapter, we will be able to determine the profile of the household that faces higher than average inflation. Is it a lower-income young family living in a village or rich pensioner who lives in a city?

6.1 Preliminary analysis

Before a regression analysis can be carried out, we have to understand the data and consider all possible drawbacks of our dataset. Subsection 6.1.1 states the hypotheses that we test, Subsection 6.1.2 describes the data, Subsection 6.1.3 summarizes possible problems of the survey data and in Subsection 6.1.4 we list variables with their expected effects on inflation.

6.1.1 Hypotheses

Qualitative analysis showed that some groups of household have inflation consistently higher/lower than the average. Results from Chapter 5 help us

to state the null hypothesis and to quantify the effects of particular household characteristics in regression analysis.

Hypothesis #1: Inflation rates of different groups of households differ significantly.

Hypothesis #2: Pensioners living alone in a city with low-income are the group with the highest inflation.

This chapter aims to analyse the hypotheses quantitatively, discuss the results and compare them with our findings from the qualitative part.

6.1.2 Data

The origin of the data that we use in the empirical analysis is described in detail in Chapter 3. The main dataset used in this study is the Czech Household Budget Survey. We have matched this dataset with third level price indices constructed by CZSO and managed to calculate specific household inflation rates, as defined in Equation 4.5.

In the HBS, around 3,000 households are interviewed every year and purposive quota sampling is used for selection of the sample. The authors of the survey try to make it as similar to the random sample as possible, but the results drawn from the survey might be biased or misleading.[1] As the HBS data are used in many Czech studies, we believe that the assumption of random sampling has not been violated and we use the data without any further restrictions.[2]

We have 71,837 observations for the years 1990–2012. The number of questioned households is not the same every year, which leads to the problem of unbalanced panel data, discussed in Subsection 6.2.2.

6.1.3 Quality of Survey data

The data we use in our analysis are survey data that can be subject to a few problems. Cameron & Trivedi (2005) describe potential issues with the survey data and we follow the structure of their work and discuss whether we are concerned by a particular problem or not.

[1] Quality of the survey data is further discussed in the next subsection.
[2] Other drawbacks of the dataset have already been discussed in Chapter 3.

Attrition bias

Usually, households are not part of the survey for the whole investigated period 1990–2012, but they quit after s few years or they are replaced by another household which better reflects characteristics needed to fill the quotas.[3] Incomplete participation can lead to attrition bias, which originates if the units composing the dataset do not leave randomly. Such bias is present in part of the households in our sample, some households are replaced to correspond to the quotas. The authors of the survey try to simulate a random sample in every period, so there is no reason to treat the data as non-randomly sampled.

Problem of survey non-response

Problem of non-response arises if the relationship between the refusal to answer a question and the characteristic of a household is systemic. This is not our case as a household either refuses to participate in the whole survey (in this case it is replaced by another household with the same characteristics and it does not affect the survey) or fills in the whole questionnaire.

Missing and mismeasured data

Cameron & Trivedi (2005) stress that missing data can lead to biases similar to the selection bias if the data are missing in a systemic way (an example can be the unwillingness of high-income groups to answer questions related to income). We do not have any missing data for variables of interest in our dataset, therefore we do not have to deal with the problem of missing data.

On the other hand, measurement error can be an issue as households tend to underestimate some expenses (e.g. alcohol or cigarettes).[4] We believe that the under/overestimation is not very high as the survey is very detailed and questioned households have to divide their total expenses between many groups of goods and services.[5] Moreover, the questionnaire is anonymous, so there is less intent to distort the results.

[3] Quotas are set based on the sample survey of Living Conditions, which is a national modification of the European survey: European Union - Statistics on Income and Living Conditions (EU-SILC). This survey is carried out by random selection in order to obtain representative data on the level and structure of income and basic socio-demographic characteristics of households.

[4] This fact was already mentioned in **??** together with other drawbacks.

[5] If they report lower alcohol expense, they have to increase expenses elsewhere, which is improbable.

6.1.4 Variables and their expected effects

In this section, selected variables and their expected effects on the level of household inflation rate are summarized. The expectations of the sign of their effects on household specific inflation are based mostly on the qualitative analysis done in Chapter 5 and we do not provide an explanation of the reasons behind our expectations in this subsection, the reasons behind the effects follows in the commentary of regression results.

As to our knowledge, no studies on this topic have conducted regression on microeconomic determinants of inflation. The choice of variables used in this thesis has been guided by observation of the data, by the qualitative analysis in Chapter 5 and by qualitative analyses done by researchers from other countries. We added to the explanatory variables another two that relate to the ownership of a summer house and the total floor area of the flat/house to examine their possible impact on households' inflation.[6] The variables are described below and their descriptive statistics are in Table 6.1 together with the expected sign of their effect on household inflation rate.

Table 6.1: Summary of variables

					Expectations	
Variable	Mean	SD	Min	Max	Sign	Significant
Income decile	5.50	2.87	1	10	-	yes
Pensioners	0.32	0.61	0	3	+	yes
Commune size	6.03	2.59	1	9	+	yes
Household size	2.65	1.24	1	11	-	yes
Children	0.89	1.02	0	9		no
Economically active	1.25	0.82	0	5.42	-	yes
Floor area	75.48	30.82	8	517		no
Summer house	0.11	0.32	0	2		no

Source: Household Budget Survey

- Income deciles

We have constructed an income decile for each year separately based on the net income of a household per month. Net income is defined for the purpose

[6] We have selected only two characteristics from many available in the data - one from dummies that relates to the ownership and one characteristic that was not examined in Chapter 5 - total floor area. The dataset is very detailed, it could be subject to further examination.

of the survey as gross income minus income taxes, compulsory insurance and withdrawn loans and deposits. We have ten deciles, 1 being the lowest income decile and 10 being the highest income decile. Section 5.2 has shown that rich households have lower inflation than poor households. If this is not caused by any other fact, we expect the effect of income on inflation to be negative.

- Number of pensioners

In this quantitative part, we use the number of pensioners in a household as an explanatory variable, we do not consider whether pensioners live in a household alone or with someone else.[7] There is either no pensioner, one, two or at maximum, three pensioners in a household. We expect households with pensioners to have higher inflation, therefore the effect of this variable could be positive if the difference is not caused by another variable.

- Commune size

This variable stands for the size of a commune that a household lives in. Table 6.2 describes the values that the variable takes. The average household is in group 6. The expected effect of the size of the commune on inflation is positive, the bigger a commune in which a household lives, the higher the inflation.

Table 6.2: Commune size

Value	Description	Value	Description
1	<499 inhabitants	6	10,000–19,999
2	500–999	7	20,000–49,999
3	1,000–1,999	8	50,000–99,999
4	1,999–4,999	9	>100,000 inhabitants
5	5,000–9,999		

Source: Household Budget Survey guidebook

- Household size - number of persons

[7] Unlike in the qualitative part of this work where we divided households into two groups - households of pensioners only and others.

This variable captures the total number of persons in a household. On average there are 2.65 persons in a household, but it ranges from 1 to 11. As in Section 5.4, we expect household size to have a significant effect on inflation. The effect should be negative as the group of households of only one person had significantly higher inflation than households of two or more people.

- Number of children

The number of children ranges from 0 to 9, the mean is 0.89. The difference between families with and without children in Section 5.5 was not very big, so we expect the effect on inflation to be insignificant.

- Number of economically active persons

The number of economically active members ranges from 0 to 5.42.[8] We expect the sign of the effect on inflation to be negative as more people working means a higher income and larger their choice of goods and services. If the price of the cheapest bread increases, they can substitute it for more expensive bread whose price has not change. We believe that this variable might be correlated with income, it is further discussed in Subsection 6.2.3.

- Total floor area of the flat/house

We have included two new variables that were not analysed in the qualitative part and could have an effect on inflation. Total floor area is defined as the total floor area in square meters (toilet, bathroom, pantry and closet are excluded). The average is 75.48 square meters, the minimum is 8 and the maximum is 517. We do not expect this variable to be significant.

- Summer house

The last possible variable of interest is the ownership of a summer house. Households in our sample either do not have a summer or have one or two. The average is only 0.11 and we again do not expect it to be significant and it might be positively correlated with the income decile (the higher the income, the higher the probability of having a summer house).

[8] The survey takes into account part-time workers, that is why the maximum number does not need to be an integer.

6.2 Methodology and testing

The general methodology that is used in this study is based mostly on Wooldridge (2002a), Wooldridge (2002b), Andress et al. (2003) and Cameron & Trivedi (2005). We specify the initial model, choose the right estimator and perform robustness testing to be able to run valid inference in Section 6.3. The selection of the right estimator follows the methodology of Park (2011) and testing follows the structure of Torres (2013).

6.2.1 Initial model specification

We based the selection of variables on the analysis from the previous chapter and other studies that conducted similar analysis as in Chapter 5. As far as we know, no other study conducted economic regression on microeconomic determinants of inflation rate.

$$\begin{aligned} inflation_{i,t} = \beta_0 + \beta_1 decile_{i,t} + \beta_2 pensioners_{i,t} + \beta_3 communesize_{i,t} + \\ + \beta_4 householdsize_{i,t} + \beta_5 children_{i,t} + \beta_6 econactive_{i,t} + \\ + \beta_7 summerhouse_{i,t} + \beta_8 floorarea_{i,t} + \alpha_i + e_{i,t} \end{aligned} \quad (6.1)$$

where β_i is a coefficient of our explanatory variables, α_i is the fixed effect (unobservable household specific effect) and $e_{i,t}$ is the error term.

There is a time dimension also for the variables that seem to be time-invariant, because we have also found some within variation for these variables. An example is the decile that is computed separately for every year, therefore one household can be in a different decile in two consecutive years. The same holds for households' characteristics - number of children or pensioners can change within one household, also commune size changes if a family moves from a city to the countryside. Wooldridge (2002b) describes this problem and states that such variables can be handled as time-varying.[9]

[9] "In panel data analysis the term *time-varying explanatory variables* means that each element of $x_{i,t}$ varies over time for *some* cross section units. Often there are elements of $x_{i,t}$ that are constant across time for a subset of the cross-sections.", Wooldridge (2002b), p. 301.

6.2.2 Panel Data vs Pooled Cross Sections Data

We have data with two dimensions - unit and time, cross sections for different periods. First, we need to state the nature of our data to be able to distinguish if we have independently pooled cross sections or panel data. A pooled cross section dataset is random sample of individual cross sections, they are obtained by collecting observations from a large population independently of each other at different periods. The probability that the same unit will be questioned more than once is very low as the dataset is collected from a large population. Panel data, on the other hand, follows the same units in every period and therefore, it is more difficult to collect the dataset.

Our dataset is a bit specific. It is not a random sample, but it is constructed in a way that is tries to make it as similar to a random sample as possible. Types of households are selected by quota sampling, but households that correspond to these characteristics are then chosen randomly. According to the needs of the authors of the survey, some households are questioned for many consecutive years, while others are replaced after one year.

After discussion with experts from the Institute of Economic Studies, the dataset is treated as unbalanced panel data. It cannot be classified as pooled cross sections as our data violates the main assumption - some units are questioned for more years, which violates the independence of cross sections. The fact that the panel is unbalanced does not pose a significant problem as Stata software can handle such a dataset.

6.2.3 Correlation matrix

The measure of interdependence of variables is shown in Table 6.3. The correlation matrix suggests there might be multicollinearity in our data and the inclusion of two correlated variables would lead to an upward bias of standard errors. Connolly (2007) defines a strong correlation of two variables as a correlation coefficient higher than 0.6. This rule of thumb helps us in the decision about inclusion or exclusion of variables.

The variables number of children and household size have a correlation coefficient of 0.89, the relationship is intuitive - the more children in a household, the bigger is its size. We drop the variable that stands for the number of

children because it has broader possibilities of interpretation.[10]

There are other variables with a correlation above 0.6 - the number of economically active persons is highly correlated with the income decile (positive correlation) and the number of pensioners (negative correlation). The relationship is again intuitive as more economically active members of a household increase the total income, which moves the household into a higher income decile. The number of working members is also correlated with the number of pensioners, households of pensioners do not usually have any economically working members. We do not include the number of economically active persons in regression as it is correlated with another two variables and there would not be a big value added.

Table 6.3: Correlation matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Decile	1.00							
(2) Pensioners	-0.29	1.00						
(3) Commune size	-0.01	0.01	1.00					
(4) Household size	0.45	-0.34	-0.12	1.00				
(5) Children	0.22	-0.41	-0.08	0.89	1.00			
(6) Econ. active	0.68	-0.64	-0.07	0.49	0.26	1.00		
(7) Summer house	0.14	0.11	0.21	-0.03	-0.10	0.05	1.00	
(8) Floor area	0.30	-0.06	-0.31	0.31	0.20	0.27	-0.03	1.00

Source: Author's calculation

6.2.4 Choice of the estimation method

Panel data can be estimated with different methods, the basic methods are the Pooled OLS estimator, the Fixed effects estimator (FE) or the Random effects estimator (RE). Three pairwise tests help us distinguish which method is the most suitable. We follow the selection method summarized in Park (2011).

F-test: Pooled OLS vs FE estimator

The F-test tests the null hypothesis that the observed and unobserved fixed effects α_i from Equation 6.1 are equal to zero, i.e. they are equal across all cross sections.

[10] Number of children is limited to families only, while household size is a characteristic of every household.

$$H_0 : \alpha_i = 0 \quad \forall\ i \tag{6.2}$$

The value of our F statistic is F(24,440; 47,390) = 1.77, its critical value at a 1% significance level is 1.03, the p-value is 0.0000. We reject the null hypothesis because the calculated F statistic is greater than the critical value for given degrees of freedom. Rejection of the null hypothesis means that we should favour the FE estimator as fixed effects are not jointly equal to zero.

LM test: Pooled OLS vs RE etimator

The Breusch-Pagan Lagrange multiplier test helps us to decide between Random effects regression and a simple OLS regression.[11] The null hypothesis is that the variance across entities is zero. Formally:

$$H_0 : \sigma_\alpha^2 = 0 \tag{6.3}$$

The reported p-value is 0.0000, therefore it is small enough to reject the null hypothesis at a 1% significance level and we conclude that there is a random effect within the data. The use of the Random effects estimator is therefore more suitable than pooled OLS.

Hausman specification test: FE vs RE estimator

Based on the results above, we know that the data should be treated as panel data and both fixed and random effects are present. The Hausman specification test helps us to decide between a Fixed or Random effects model. The null hypothesis states that coefficients estimated by the *efficient* RE estimator are the same as the ones estimated by the *consistent* FE estimator. RE is preferred if the null hypothesis is not rejected and FE otherwise.

$$H_0 : \beta_i^{FE} \approx \beta_i^{RE} \quad \forall\ i \tag{6.4}$$

We reject the null hypothesis because the estimated value of $\chi^2(6)$ is 202.71, while the critical value of $\chi^2(6)$ at a 1% significance level is 16.81 (the p-value 0.0000). Based on the Hausman specification test, we use the Fixed effects estimation method in our analysis since Random effects estimation would be inconsistent.

[11] The test was developed by Breusch & Pagan (1980) and modified by Baltagi & Li (1990) to allow for unbalanced panel data. Stata command already has the modification incorporated.

6.2.5 Robustness testing

Before we proceed to empirical testing, several robustness checks have to be carried out to obtain valid inference later. For the purpose of this subsection, we use the updated model, where two correlated variables are excluded (number of children and number of economically active people in a household). Torres (2013) lists, in the diagnostics section, all the tests that should be run and we follow his structure.

Testing for time effects

The first test shows us if time effects should be controlled when running the FE model. It tests jointly whether all dummies for years are equal to zero. If we reject the null hypothesis, we need to include time effects.

$$H_0 : \gamma_t = 0 \quad \forall\ t \tag{6.5}$$

The value of the F statistic is F(22; 47,368) = 42,896, its critical value is 1.83 at a 1% significance level and the p-value is 0.0000, therefore we reject the null hypothesis that coefficients of time effects are jointly equal to zero and we need to include time effects in our analysis.

Unit root test

Wooldridge (2002a) states that regressing variables integrated of order one may lead to a spurious regression. This happens when unrelated variables show statistically significant dependence without any causal dependence. Under the null hypothesis, the variable is non-stationary, formally stated:

$$H_0 : \textit{All panels contain unit roots.} \tag{6.6}$$

There exists many unit root tests, we test the presence of non-stationarity by the Augmented Dickey-Fuller test, because it can be used in case of unbalanced panel data. We perform the test on all our variables and we reject the null hypothesis at 1% significance level (the p-value 0.0000) for all of them. We do not have to deal with the problem of non-stationarity.

Test for serial correlation

We use Wooldridge (2002b) test for autocorrelation in panel data that is incorporated in Stata software. The null hypothesis is defined as:

$$H_0 : \textit{No first-order autocorrelation.} \tag{6.7}$$

The result from the test is F(1; 7,390) = 6.784 and the critical value at a 1% significance level is 6.64 (the p-value 0.0092), so we reject the null hypothesis and conclude that the data has first-order autocorrelation and we need to correct for it.

Heteroskedasticity test

One of the assumptions is that unobservable error has constant variance (assumption of homoskedasticity). Andress et al. (2003) state that if the assumption of homoskedastic error terms is violated, FE estimates are inefficient, estimated error terms are biased and significance tests may produce incorrect conclusions. The null hypothesis of the test is defined as:

$$H_0 : \sigma_i^2 = \sigma^2 \quad \forall\ i \tag{6.8}$$

The p-value is 0.000, so we reject the null hypothesis on a 1% significance level (the test statistic is very high: $\chi^2(24,441) = 6.3\text{e}{+37}$) and we will have to adjust the inference by robust standard errors.

Multicollinearity

We have already investigated the correlation matrix in Table 6.3 and we excluded two variables. A formal test did not find any collinear variables, so we do not have to deal with it.

6.2.6 Model re-specification and correction

Before proceeding to regression analysis and results, we specify a new model based on the results of the correlation matrix in Table 6.3. We decided to drop two variables with a correlation higher than 0.6 - number of economically active members of a household and number of children. Based on the test for time effects in Section 6.2.5 we need to include a dummy variable for every year.[12] The re-specified model that is further estimated is defined as:

$$\begin{aligned} inflation_{i,t} = \beta_0 + \beta_1 decile_{i,t} + \beta_2 pensioners_{i,t} + \beta_3 communesize_{i,t} + \\ + \beta_4 householdsize_{i,t} + \beta_5 summerhouse_{i,t} + \beta_6 floorarea_{i,t} + \\ + \gamma_1 t_{1991} + ... + \gamma_{22} t_{2012} + \alpha_i + e_{i,t} \end{aligned} \tag{6.9}$$

[12] To avoid perfect multicollinearity, we implement only 22 dummies for years, even though we have 23 years.

where β_i are coefficients of our explanatory variables, γ_i are coefficients of the year dummy variables and $e_{i,t}$ is the error term.

During our testing, we found both heteroskedasticity and autocorrelation present in the data. Arellano (1987) first proposed the method that is known as clustered standard errors.[13] With the use of this method, standard errors are robust to any kind of serial correlation or heteroskedasticity. We use this method as it is suitable for Fixed effects estimation and clustered standard errors are incorporated in Stata software.

6.3 Empirical results

FE estimation with time effects and clustered standard errors yielded the results summarized in Table 6.4. All four variables that are added to the model based on results from Chapter 5 have significant coefficients, but coefficients of summer house and floor area are not significant. This is in line with the first hypothesis[14] and our expectations as stated in Table 6.1. All coefficients of significant variables have expected sign, which complies with our second hypothesis about the effect of variables on inflation of a household.[15]

6.3.1 Discussion of results for significant variables

- Income deciles

Level of income has a negative effect on inflation and the coefficient is significant on a 1% significance level. The effect seems to be relatively low, but we have ten deciles and every move of a household to a higher decile decreases the inflation by almost 0.1 of a percentage point. The difference between two households that would differ only in level of income would be 0.88 of a percentage point (rich households in decile 10 have lower inflation than households in decile 1). A difference of 0.88 of a percentage point is very big since the introduction of inflation targeting, inflation is kept mostly below 5%.

The reason behind this could be that low-income households are forced to spend a higher share of their expenses on necessities such as food and rent,

[13] For further description of the method, please refer to Wooldridge (2002b) or Vogelsang (2012).

[14] *Hypothesis #1: Inflation rates of different groups of households differ significantly.*

[15] *Hypothesis #2: Pensioners living alone in a city with low-income are the group with the highest inflation.*

Table 6.4: FE estimation method with clustered standard errors

Variable	Coefficient	Std. Errors	P-value
Decile	-0.0009766**	(0.0000861)	0.000
Pensioners	0.0010359*	(0.0004409)	0.019
Communesize	0.0011703**	(0.0001143)	0.000
Householdsize	-0.0004745*	(0.0002323)	0.041
Summerhouse	0.0010769	(0.0007328)	0.142
Floorarea	0.0000033	(0.0000078)	0.675
1991.year	0.436**	(0.002)	0.000
1992.year	-0.022**	(0.001)	0.000
1993.year	0.015**	(0.001)	0.000
1994.year	-0.069**	(0.001)	0.000
1995.year	-0.079**	(0.001)	0.000
1996.year	-0.090**	(0.001)	0.000
1997.year	-0.089**	(0.001)	0.000
1998.year	-0.054**	(0.001)	0.000
1999.year	-0.164**	(0.001)	0.000
2000.year	-0.130**	(0.001)	0.000
2001.year	-0.127**	(0.001)	0.000
2002.year	-0.172**	(0.001)	0.000
2003.year	-0.190**	(0.001)	0.000
2004.year	-0.157**	(0.001)	0.000
2005.year	-0.163**	(0.001)	0.000
2006.year	-0.152**	(0.001)	0.000
2007.year	-0.151**	(0.001)	0.000
2008.year	-0.102**	(0.001)	0.000
2009.year	-0.160**	(0.001)	0.000
2010.year	-0.161**	(0.001)	0.000
2011.year	-0.146**	(0.001)	0.000
2012.year	-0.131**	(0.001)	0.000
Intercept	0.181**	(0.001)	0.000
Observations		71 837	
R^2		0.9599	
F(28,24440)		10,305.33	0.0000

Note: * p <0.05, ** p <0.01

Source: Author's calculations

they cannot substitute these expenses if their prices rise and the prices of these expenditures usually increase the most. Households with a high income have a better choice of goods and services and if the price of the cheapest bread increases they can substitute it by more expensive bread whose price did not change that much. Results of the regression imply that high inflation further increases inequality between rich and poor and it would be an interesting topic for another study.

- Number of pensioners

The variable representing the effect of number of pensioners on inflation has a positive effect significant on a 5% significance level and every pensioner increases a household inflation by 0.1 of a percentage point. The effect is lower than we would expect it to be after the qualitative part, which means that other factors such as the income decile have higher effects. Pensioners are usually in the low-income decile, which implies higher inflation and the fact that they are retired does not have a very big effect. Hobijn & Lagakos (2003) explained the higher inflation of pensioners by rising price of health care. Another reason is that pensioners mostly spend on goods and services whose prices increase the most and cannot be substituted (rent, food and drugs).

Pensioners are a vulnerable group of the population and their pensions are indexed according to the inflation rate and also according to the growth rate of nominal wages. According to our results, pensioners have a higher inflation rate than the official one for the whole population which is used to increase their pensions. Their pensions should be increased more if policy makers considered group-specific inflation for pensioners.

Interestingly, CZSO publishes overall consumer price indices for the whole population ("official" inflation rate), for Prague and for pensioners. Even though the inflation rate for pensioners is known and the differences are considerable, the official inflation rate is used for the indexation of pensions. Although a different approach to the implementation of fiscal measures is usually problematic, the indexation of pensions only applies to pensioners and it would not cause fragmentation of fiscal policy.

- Commune size

The coefficient of the variable commune size has positive effect significant on a 1% significance level. Its effect is quite big if we consider that the variable

takes values from 1 to 9. Holding other characteristics equal, a household in a village with less than 499 inhabitants has a lower inflation by 0.94 of a percentage point than household in a city with more than 100,000 inhabitants. We found that much of the higher inflation for urban area is caused by increases in the cost of accommodation, mainly rent.

- Household size - number of persons

The effect of the size of a household also confirms the expected negative sign and the coefficient is significant on a 5% level. Household size ranges from 1 to 11 and if we compare the extremes - a single person household has higher inflation by 0.47 of a percentage point than household with eleven members. A single person household uses a higher proportion of its expenditures on goods and services that are paid only once regardless of the size of a household (rent, fuel cost or flat equipment) and cannot be substituted.

6.3.2 Groups of households with the extreme inflation

We now compare the difference in inflation rate between the lowest and highest inflation household profile based on our regression results. We compare inflation for the year 2012, the last year in the examined period of two households - three pensioners in one household with the lowest income (decile 1) living in a city with more than 100,000 inhabitants and a rich family (decile 10) with nine children, which lives in a small village (less than 499 inhabitants).

A defined household of three pensioners would have inflation in 2012 equal to 6.15%, while a big family would face inflation of 3.62%. These households are at the very end range of the inflation distribution, but they might exist and face the inflation rate that differs by more than two and half percentage points. The mean inflation for that year was 4.93%, so only slightly above the average of our two extremes.

6.3.3 Further research opportunities

This study does not aim to draw any implications or propositions for fiscal policy, because it would require discussion in a large extent, but it is an open topic for further research. We summarize several possible ways the research could proceed in a few paragraphs below.

We have identified groups with higher or lower inflation, it could be interesting to further analyse reasons behind it and look closer at the consumer

basket of specific groups. Hait & Janský (2014) analysed contributors to the inflation rate for pensioners and low-income households, it would be interesting to examine other groups and decompose contributors not only at the first level of COICOP categories, but for specific goods and services. Further research in that field could build upon our study and deeply focus on causes and propose possible solutions.

We have already mentioned that pensions are indexed according to the increase of the nominal wage and according to the official rate of inflation. Part of the further analysis described above could be the quantification of the differences in pension increases if indexation were based on pensioner's specific inflation.[16]

Crawford & Smith (2002) analysed the effect of ignoring different inflation on measuring income inequality. They found out that the use of the official inflation for everyone can over/understate the annual growth rate in inequality by six percentage points. The same analysis for the Czech Republic could yield interesting results as differences between different groups are bigger and more persistent than in the UK.

These are only a few examples of possible research, the area is rather unexplored and there are many policy research opportunities. We believe that this study will serve as a springboard for further research.

[16] We do not include such quantification as it would need a discussion of policy implications and it would not be consistent with our topic.

Chapter 7

Conclusion

This thesis investigated the differences in inflation between Czech households for the period 1990–2012. Using the detailed data from the Household Budget Survey, we calculated inflation for every household in the dataset, compared it with the official average inflation and constructed subgroup inflation rates for households with similar spending patterns. The analysis showed that the differences in inflation of different households are bigger and more persistent than in other countries and the level of inflation for selected subgroups of the population should be monitored if not used in policy making.

We provide several interesting findings in this study. When examining inflation rate representativeness, we have found that on average only 69% of households have inflation close to the mean inflation, where "closeness" was defined as maximum from i) percentage of households within 25% around the mean inflation and ii) percentage of households within one percentage point around the mean inflation. Our definition of closeness is more suitable for years with both high and low inflation and it showed that the variation is high and it is reasonable to speak about group specific price indices.

The aim of the qualitative part was the identification of groups with consistently high or low inflation, which was done by the construction of subgroup inflation rates for groups of households with similar patterns selected based on the previous studies. We have constructed inflation rates for households with different incomes, households of pensioners and of a single person, households from urban or rural areas and families with children. We showed that socially-weaker groups such as low-income households and pensioners tend to have consistently higher inflation than the average population, other groups with higher inflation are single households and households living in bigger cities.

The biggest difference was in 2009 between low- and high-income households, when it reached 3.67 percentage points. This was an extreme difference, but average differences between groups were close to one percentage point, which is a big difference in case of inflation. After the analysis, we might say that inflation increases the inequality between rich and poor and it would be advisable to consider compensations to these vulnerable groups of population.

The main contribution of this thesis is the unbalanced panel data regression in the quantitative part. We were the first to introduce panel data regression in that area and it allowed us to examine all the effects of household characteristics at once to avoid the possibility that higher inflation of low-income households is caused by other reasons than income. Out of all considered estimators, testing helped us to choose the fixed effects estimator with time effects and clustered standard errors.

In general, our empirical results cannot reject our hypotheses, main variables are significant on a 5% significance level. The effect of income and household size proved to be negative as expected. According to our results household in decile 1 would have of a percentage point higher inflation than the one at decile 10. Our results suggest that rich households can substitute for the goods whose price increased the most. The effect of household size is lower - households with four members would have a higher inflation by 1.4 percentage points than a single person household. Bigger households use a smaller part of their expenses on rent, fuel costs and other expenses that are paid only once regardless of the size of the household, where the increase of prices is usually the highest. The number of pensioners and size of a commune, on the other hand, have positive effect on inflation. The size of a commune has very big effect - a household in a village would have an inflation that is 0.94 of a percentage point lower inflation than inflation of a household living in a big city, the reason is the increasing price of rent. The effect of pensioners is lower - every pensioner increases inflation by 0.1 of a percentage point. Pensioners spend mostly on food, fuel, rent and drugs - goods that cannot be substituted for and their price increases fast.

At the end we are able to identify households with characteristics leading to the highest inflation or the lowest inflation. A household of low-income pensioners living in a big city is the household with the highest inflation, while rich families with many children living in a small village have the lowest inflation. Surprisingly, this difference can be more than 2.5 percentage points, which is a lot if we consider that the official inflation has been below 5% in the last years.

The combination of results suggests that the CPI is a reasonable measure of inflation, but we have to be careful when stating that it represents inflation for the whole population. This thesis proved that the differences in inflation of Czech households are significant and should be monitored. Construction of subgroup indices for monitoring purposes would not cause any additional costs as the CZSO already has the information about all groups of households from the Household Budget Survey. Even though the differences are large, general introduction of different price indices would cause distortion and it could bring more negative than positive things. Nevertheless, policy makers should consider the application of subgroup price indices in the cases where the policy measures relate only to one group, which has inflation higher or lower than average. An example is indexation of pensions, where the use of the pensioner's subgroup consumer price index would not cause any disruption of the policy as pensioners are the only group concerned by indexation of pensions.

Further extension of this work could focus more on the reasons for the differences in inflation between groups of households. We have measured the size of the difference and proposed main reasons, but a more detailed analysis would contribute to a better understanding of the inflation differentials. We believe that this thesis will open the discussions on the topic in the Czech Republic and it will be subject of further research.

Bibliography

Andress, H.-J., K. Golsch, & A. W. Schmidt (2013): "*Applied panel data analysis for economic and social surveys.*" Springer Science & Business Media.

Arellano, M. (1987): "*Computing Robust Standard Errors for Within Group Estimators.*" Oxford bulletin of Economics and Statistics 49(4): pp. 431–434.

Arrow, K. J. (1958): "*The measurement of price changes.*" Joint Economic Committee, US Congress, Washington, DC Reprinted in Collected Works of Kenneth J. Arrow.

Artsev, Y., V. Roshal, & Y. Finkel (2006): "*Consumer Price Indices–Measuring Across Households.*" Ninth Ottawa Group Meeting on Prices, London.

Baltagi, B. H. & Q. Li (1990): "*A Lagrange multiplier test for the error components model with incomplete panels.*" Econometric Reviews 9(1): pp. 103–107.

Berndt, E. R., I. M. Cockburn, D. L. Cocks, A. M. Epstein, & Z. Griliches (1998): "*Is price inflation different for the elderly? An empirical analysis of prescription drugs.*" Forum for Health Economics & Policy 1(1).

BLS (1997, updated 06/2007): "*Handbook of Methods.*" US Department of Labor, Chapter 17.

Boskin, M. J., E. R. Dulberger, Z. Griliches, R. J. Gordon, & D. Jorgensen (1996): "*Toward a more accurate measure of the cost of living: final report to the Senate Finance Committee for the advisory commision to study the consumer price index.*" Final report.

Breusch, T. S. & A. R. Pagan (1980): "*The Lagrange multiplier test and its applications to model specification in econometrics.*" The Review of Economic Studies pp. 239–253.

Cage, R., J. Greenlees, & P. Jackman (2003): "*Introducing the chained consumer price index.*" International Working Group on Price Indices (Ottawa Group): Proceedings of the Seventh Meeting pp. 213–246.

Cameron, A. C. & P. K. Trivedi (2005): "*Microeconometrics: methods and applications.*" Cambridge university press.

Connolly, P. (2007): "*Quantitative data analysis in education: A critical introduction using SPSS.*" Routledge.

Crawford, I. & Z. Smith (2002): "*Distributional aspects of inflation.*" Institute for Fiscal Studies.

CZSO (1989,1990): "*Statistical Yearbook of the Czech Republic.*" Years 1989, 1990.

CZSO (2015a): "Indexy spotřebitelských cen - inflace březen 2015." https://www.czso.cz/csu/czso/cri/indexy-spotrebitelskych-cen-inflace-brezen-2015 *[Online; accessed 12-May-2015]*.

CZSO (2015b): "*Czech Statistical Office.*" https://www.czso.cz/ *[Online; accessed 12-May-2015]*.

DeMilner, L. E. (1981): "*Indexing with the consumer price index: problems and alternatives.*" The Office.

Eurostat (2004): "*Household Budget Survey in the Candidate Countries. Methodological Analysis 2003.*" Luxembourg: Office for Official Publications of the European Communities.

Garner, T. I., D. S. Johnson, & M. F. Kokoski (1996): "*An Experimental Consumer Price Index for the Poor.*" Monthly Lab. Rev. 119: p. 32.

Gordon, R. J. (2000): "*The Boskin Commission report and its aftermath.*" National bureau of economic research.

Gordon, R. J. (2006): "*The Boskin Commission report: A retrospective one decade later.*" National Bureau of Economic Research.

Hait, P. & P. Janský (2014): "*Inflation Differentials among Czech Households.*" CERGE-EI.

Hanousek, J. & R. K. Filer (2001): "*Evaluating imperfections and biases in price indexes during transition.*" CERGE-EI Working Paper Series (186).

Hobijn, B. & D. Lagakos (2005): "*Inflation inequality in the United States.*" Review of income and Wealth 51(4): pp. 581–606.

Idson, T. & C. Miller (1999): "*Calculating a price index for families with children: implications for measuring trends in child poverty rates.*" Review of Income and Wealth 45(2): pp. 217–233.

Leicester, A., C. O'Dea, & Z. Oldfield (2008): "*The inflation experience of older households.*" Institute for Fiscal Studies.

Leszczynska, A. & A. Halka (2012): "*What does the Consumer Price Index Measure? Bias Estimates for Poland.*" EcoMod.

Lieu, P.-T., C. Chang, & J.-r. Chang (2004): "*Inflation rate variations across household: empirical evidence from Taiwan.*" International Journal of Business(**1**).

Murphy, E. & E. Garvey (2004): "*A consumer price index for low-income households in Ireland (1989-2001).*" Combat Poverty Agency, Working Paper 4(03).

Neves, P. D. & L. M. Sarmento (1997): "*The Substitution bias of the consumer price index.*" Banco de Portugal, Economic Bulletin pp. 25–33.

OECD (2015): "*Main indicators: Czech Republic: Consumer Price Index.*" http://stats.oecd.org/mei/default.asp?lang=e&subject=8&country=CZE *[Online; accessed 12-May-2015]*.

Oosthuizen, M. (2007): "*Consumer Price inflation across the income distribution in South Africa.*" Development Policy Research Unit, University of Cape Town.

Park, H. M. (2011): "*Practical guides to panel data modeling: a step-by-step analysis using Stata.*" Public Management and Policy Analysis Program, Graduate School of International Relations, International University of Japan.

Podpiera, J. (2003): "*Does CPI Approximate Cost-of-Living?: Evidence from the Czech Republic.*" CERGE-EI.

Prais, S. J. (1959): "*Whose cost of living?*" The review of economic studies, JSTOR pp. 126–134.

Sojka, M. (1994): "*The Transformation of the Czech Economy-Present and Future Developments.*" Czech Sociological Review, JSTOR pp. 35–51.

Sugema, I., M. I. Irfany, A. Holis, & T. Bakhtiar (2010): "*Consumer Price Index for the Poor (CPI-P): An Empirical Analysis of Indonesia.*" International Research Journal of Finance and Economics 58.

Torres, R. (2013): "*Panel Data analysis fixed and random effects using Stata 10. x.*" Data & Statistical Services.

UZIS (2015): "*Institute of Health information and Statistics of the Czech Republic.*" https://http://www.uzis.cz/ *[Online; accessed 12-May-2015].*

Vogelsang, T. J. (2012): "*Heteroskedasticity, autocorrelation, and spatial correlation robust inference in linear panel models with fixed-effects.*" Journal of Econometrics, Elsevier 166(2): pp. 303–319.

Wooldridge, J. M. (2002a): "*Econometric analysis of cross section and panel data.*" MIT press.

Wooldridge, J. M. (2002b): "*Introductory econometrics: A modern approach.*" Cengage Learning.

Appendix A

COICOP classification

Table A.1: COICOP third level classes

Classes marked with * are not used in the Czech Republic. Either they are not part of the survey of the CZSO (023 Narcotics, 0714 Animal drawn vehicals, 122 Prostitution or 1251 Life Insurance) or they are not relevant for the Czech Republic (0734 Passenger transport by sea and inland waterway). From the total 176 classes, the CZSO reports 162 classes.

Class number	Class name
01	Food and non-alcoholic beverages
011	Food
0111	Bread and cereals
0112	Meat
0113	Fish and seafood
0114	Milk, cheese and eggs
0115	Oils and fats
0116	Fruit
0117	Vegetables
0118	Sugar, jam, honey, chocolate and confectionery
0119	Food products n.e.c.
012	Non-alcoholic beverages
0121	Coffee, tea and cocoa
0122	Mineral waters, soft drinks, fruit and vegetable juices
02	Alcoholic beverages, tobacco and narcotics
021	Alcoholic beverages
0211	Spirits
0212	Wine
0213	Beer
022	Tobacco
0220	Tobacco

Class number	Class name
023*	Narcotics*
0230*	Narcotics*
03	Clothing and footwear
031	Clothing
0311	Clothing materials
0312	Clothing
0313	Other articles of clothing and clothing accessories
0314	Cleaning, repair and hire of clothing
032	Footwear including repair and hire
0321	Shoes and other footwear
0322	Repair and hire of footwear
04	Housing, water, electricity, gas and other fuels
041	Actual rentals for housing
0411	Actual rentals paid by tenants
0412*	Other actual rentals*
042	Imputed rentals for housing
0421	Imputed rentals of owner-occupiers
0422*	Other imputed rentals*
043	Maintenance and repair of the dwelling
0431	Materials for the maintenance and dwelling repair
0432	Services for the maintenance and dwelling repair
044	Other services relating to the housing
0441	Water supply
0442	Refuse collection
0443	Sewage removal
0444	Other services relating to the housing n.e.c.
045	Electricity, gas and other fuels
0451	Electricity
0452	Gas
0453	Liquid fuels
0454	Solid fuels
0455	Heat energy
05	Furnishings, household equipment and maintenance
051	Furniture and furnishings, carpets and floor coverings
0511	Furniture and furnishings
0512	Carpets and other floor coverings
0513	Repair of furniture, furnishings and floor coverings
052	Household textiles
0520	Household textiles
053	Household appliances including repair
0531	Major household appliances whether electric or not
0532	Small electric household appliances
0533	Repair of household appliances
054	Glassware, tableware and household utensils

Class number	Class name
0540	Glassware, tableware and household utensils
055	Tools and equipment for house and garden
0551	Major tools and equipment including repair
0552	Small tools and miscellaneous accessories and repair
053	Household appliances including repair
0531	Major household appliances whether electric or not
0532	Small electric household appliances
0533	Repair of household appliances
054	Glassware, tableware and household utensils
0540	Glassware, tableware and household utensils
055	Tools and equipment for house and garden
0551	Major tools and equipment including repair
0552	Small tools and miscellaneous accessories
056	Goods and services for household maintenance
0561	Non-durable household goods
0562	Domestic services and household services
06	Health
061	Pharmaceuticals and medical devices
0611	Pharmaceuticals
0612	Other medical products
0613	Therapeutic appliances and equipment
062	Outpatient services
0621	Outpatient medical services
0622	Dental services
0623	Other outpatient health care
0630	Hospital services
07	Transport
071	Purchase of vehicles
0711	Motor cars
0712	Motor cycles
0713	Bicycles
0714*	Animal drawn vehicles*
072	Operation of personal transport equipment
0721	Spare parts for personal transport equipment
0722	Fuels and for personal transport equipment
0723	Maintenance of personal transport equipment
0724	Other services of personal transport equipment
073	Transport services
0731	Passenger transport by railway
0732	Passenger transport by road
0733	Passenger transport by air
0734	Passenger transport by sea and inland water
0735	Combined passenger transport
0736	Other purchased transport services

Class number	Class name
08	Communication
081	Postal services
0810	Postal services
082	Telephone and telefax equipment
0820	Telephone and telefax equipment
083	Telephone and telefax services
0830	Telephone and telefax services
09	Recreation and culture
091	Audio-visual and photographic equipment
0911	Equipment for recording of sound and pictures
0912	Photographic and cinematographic equipment
0913	Information processing equipment
0914	Recording media
0915	Repair of audio-visual and photographic equipment
092	Other major durables for recreation and culture
0921*	Major durables for outdoor recreation*
0922	Musical instruments and durables for free time
0923*	Maintenance of durables for recreation and culture*
093	Other recreational equipment, gardens and pets
0931	Games, toys and hobbies
0932	Equipment for sport and outdoor recreation
0933	Gardens, plants and flowers
0934	Pets and related products
0935	Veterinary and other services for pets
094	Recreational and cultural services
0941	Recreational and sporting services
0942	Cultural services
0943*	Games of chance*
095	Newspapers, books and stationery
0951	Books
0952	Newspapers and periodicals
0953	Miscellaneous printed matter
0954	Stationery and drawing materials
096	Package holidays
0960	Package holidays
10	Education
101	Pre-primary and primary education
1010	Pre-primary and primary education
102	Secondary education
1020	Secondary education
103	Post-secondary non-tertiary education
1030	Post-secondary non-tertiary education
104	Tertiary education
1040	Tertiary education

Class number	Class name
105	Education not definable by level
1050	Education not definable by level
11	Restaurants and hotels
111	Catering services
1111	Restaurants, cafes and the like
1112	Canteens
112	Accommodation services
1120	Accommodation services
12	Miscellaneous goods and services
121	Personal care
1211	Hairdressing salons and personal grooming stuff
1212	Electric appliances for personal care
1213	Other appliances and products for personal care
122*	Prostitution*
1220*	Prostitution*
123	Personal effects n.e.c.
1231	Jewellery, clocks and watches
1232	Other personal effects
124	Social work
1240	Social protection
125	Insurance
1251*	Life insurance*
1252	Insurance connected with housing
1253	Insurance connected with health
1254	Insurance connected with transport
1255*	Other insurance*
126	Financial services
1261*	Financial services indirectly measured*
1262	Financial services
127	Other services
1270	Other services

Source: Statistical metainformation system, Czech Statistical Office

Appendix B

Results tables

Table B.1: Measure 1: Percentage of households 25% around the mean inflation rate

Year	Democratic average inflation	Lower bound M1	Upper bound M1	Households inside the range
1990	18.04%	13.53%	22.55%	91.21%
1991	61.68%	46.26%	77.10%	90.73%
1992	16.01%	12.01%	20.01%	69.77%
1993	19.70%	14.77%	24.62%	97.34%
1994	11.20%	8.40%	13.99%	76.90%
1995	10.30%	7.72%	12.87%	96.19%
1996	8.99%	6.74%	11.24%	94.50%
1997	9.34%	7.01%	11.68%	75.88%
1998	12.82%	9.61%	16.02%	66.11%
1999	1.75%	1.31%	2.19%	29.78%
2000	5.11%	3.83%	6.39%	56.01%
2001	5.45%	4.09%	6.82%	49.35%
2002	0.96%	0.72%	1.20%	15.73%
2003	-0.81%	-1.01%	-0.61%	22.56%
2004	2.41%	1.81%	3.01%	62.29%
2005	1.80%	1.35%	2.25%	37.83%
2006	2.92%	2.19%	3.65%	38.11%
2007	3.12%	2.34%	3.89%	44.00%
2008	7.96%	5.97%	9.95%	67.32%
2009	2.14%	1.61%	2.68%	16.09%
2010	2.01%	1.51%	2.51%	23.65%
2011	3.49%	2.62%	4.36%	52.79%
2012	4.93%	3.70%	6.16%	64.09%

Source: CZSO, HBS data and author's calculations

Table B.2: Measure 2: Percentage of households one percentage point around the mean inflation rate

Year	Democratic average inflation	Lower bound M2	Upper bound M2	Households inside the range
1990	18.04%	17.04%	19.04%	48.77%
1991	61.68%	60.68%	62.68%	1.96%
1992	16.01%	15.01%	17.01%	15.53%
1993	19.70%	18.70%	19.70%	36.73%
1994	11.20%	10.20%	12.20%	26.75%
1995	10.30%	9.30%	11.30%	61.19%
1996	8.99%	7.99%	9.99%	70.10%
1997	9.34%	8.34%	10.34%	37.25%
1998	12.82%	11.82%	13.82%	24.36%
1999	1.75%	0.75%	2.75%	60.56%
2000	5.11%	4.11%	6.11%	45.01%
2001	5.45%	4.45%	6.45%	37.87%
2002	0.96%	-0.04%	1.96%	58.85%
2003	-0.81%	-1.81%	0.19%	82.02%
2004	2.41%	1.41%	3.41%	83.33%
2005	1.80%	0.80%	2.80%	69.44%
2006	2.92%	1.92%	3.92%	50.34%
2007	3.12%	2.12%	4.12%	54.86%
2008	7.96%	6.96%	8.96%	37.76%
2009	2.14%	1.14%	3.14%	29.25%
2010	2.01%	1.01%	3.01%	45.12%
2011	3.49%	2.49%	4.49%	58.51%
2012	4.93%	3.93%	5.93%	55.59%

Source: CZSO, HBS data and author's calculations

Table B.3: Mean absolute deviation (MAD)

Year	Democratic average inflation	MAD
1990	18.04%	1.76
1991	61.68%	7.02
1992	16.01%	3.44
1993	19.70%	1.66
1994	11.20%	1.96
1995	10.30%	0.97
1996	8.99%	0.84
1997	9.34%	1.68
1998	12.82%	2.73
1999	1.75%	0.99
2000	5.11%	1.33
2001	5.45%	1.75
2002	0.96%	0.97
2003	-0.81%	0.59
2004	2.41%	0.66
2005	1.80%	0.85
2006	2.92%	1.25
2007	3.12%	1.08
2008	7.96%	1.70
2009	2.14%	2.10
2010	2.01%	1.49
2011	3.49%	1.08
2012	4.93%	1.19

Table B.4: Inflation rates by income deciles (1 is the lowest, 10 is the highest income)

Year	1	2	3	4	5	6	7	8	9	10
1990	17.99%	17.77%	17.90%	18.54%	18.12%	17.93%	18.19%	18.13%	18.06%	17.79%
1991	60.87%	60.67%	61.30%	62.39%	62.12%	61.31%	61.32%	62.30%	62.16%	62.37%
1992	19.46%	18.15%	17.14%	16.80%	15.63%	15.25%	15.01%	14.75%	14.22%	13.66%
1993	19.17%	19.67%	19.70%	19.83%	19.73%	19.78%	19.67%	19.73%	19.77%	19.92%
1994	11.57%	11.35%	11.19%	11.54%	11.55%	11.11%	11.26%	10.87%	10.77%	10.75%
1995	10.79%	10.64%	10.52%	10.49%	10.38%	10.50%	10.20%	10.00%	9.91%	9.53%
1996	9.54%	9.30%	9.06%	9.20%	9.14%	9.01%	8.99%	8.76%	8.57%	8.34%
1997	11.16%	10.49%	9.66%	9.68%	9.21%	8.92%	8.57%	8.77%	8.63%	8.31%
1998	16.43%	15.63%	13.89%	13.27%	12.49%	12.09%	11.61%	11.31%	11.16%	10.32%
1999	2.07%	2.13%	1.80%	1.65%	1.59%	1.69%	1.68%	1.55%	1.72%	1.63%
2000	4.95%	5.14%	5.01%	5.11%	5.31%	5.19%	5.22%	5.25%	5.03%	4.89%
2001	7.32%	6.80%	6.11%	5.76%	5.32%	5.09%	4.98%	4.68%	4.37%	4.10%
2002	1.99%	1.73%	1.29%	1.04%	0.84%	0.66%	0.61%	0.52%	0.49%	0.40%
2003	-0.54%	-0.63%	-0.78%	-0.73%	-0.83%	-0.95%	-0.90%	-0.90%	-0.92%	-0.92%
2004	2.56%	2.53%	2.52%	2.52%	2.47%	2.47%	2.36%	2.39%	2.23%	2.03%
2005	2.21%	2.20%	2.00%	2.04%	1.80%	1.70%	1.61%	1.64%	1.50%	1.32%
2006	3.76%	3.89%	3.35%	3.22%	2.89%	2.78%	2.73%	2.41%	2.33%	1.87%
2007	3.70%	3.62%	3.49%	3.27%	3.19%	3.13%	2.97%	2.85%	2.57%	2.37%
2008	9.91%	9.35%	8.79%	8.48%	8.09%	7.84%	7.25%	7.04%	6.59%	6.27%
2009	4.34%	3.96%	3.20%	2.48%	2.22%	1.62%	1.24%	0.86%	0.85%	0.67%
2010	1.92%	2.02%	2.11%	1.90%	2.10%	2.18%	2.01%	2.12%	2.01%	1.73%
2011	3.95%	3.64%	3.72%	3.37%	3.66%	3.55%	3.51%	3.50%	3.12%	2.86%
2012	5.82%	5.46%	5.27%	5.18%	4.95%	4.84%	4.79%	4.57%	4.36%	4.06%

Table B.5: Inflation rates for different subgroups

Year	Pensioners	Non-pensioners	One person	More people	Children	No children	City	Village
1990	18.46%	17.98%	17.74%	18.10%	18.02%	17.78%	17.97%	18.18%
1991	61.39%	61.73%	60.61%	61.90%	62.28%	60.55%	62.19%	60.74%
1992	18.23%	15.56%	18.57%	15.35%	15.59%	15.84%	16.85%	14.16%
1993	19.27%	19.78%	19.53%	19.73%	19.91%	19.67%	20.00%	18.92%
1994	12.29%	10.95%	12.52%	10.86%	10.96%	10.95%	12.01%	9.66%
1995	11.03%	10.12%	11.00%	10.12%	10.15%	10.03%	10.38%	10.09%
1996	9.52%	8.86%	9.57%	8.86%	8.90%	8.74%	9.15%	8.71%
1997	10.24%	9.11%	10.63%	9.01%	9.04%	9.28%	9.82%	8.24%
1998	15.01%	12.28%	15.51%	12.15%	12.17%	12.52%	13.62%	11.36%
1999	1.70%	1.76%	2.32%	1.62%	1.63%	2.10%	1.91%	1.47%
2000	5.12%	5.11%	4.94%	5.15%	5.02%	5.22%	4.82%	5.65%
2001	6.85%	5.12%	6.62%	5.18%	5.07%	5.22%	5.38%	5.59%
2002	1.47%	0.84%	1.84%	0.74%	0.77%	0.99%	1.20%	0.50%
2003	-0.51%	-0.86%	-0.37%	-0.91%	-0.97%	-0.60%	-0.61%	-1.17%
2004	2.62%	2.37%	2.56%	2.37%	2.34%	2.46%	2.37%	2.47%
2005	2.33%	1.71%	2.17%	1.71%	1.58%	1.87%	1.73%	1.93%
2006	3.85%	2.77%	3.56%	2.76%	2.63%	2.91%	2.86%	3.01%
2007	3.31%	3.08%	3.39%	3.04%	3.01%	3.18%	3.18%	3.03%
2008	9.07%	7.78%	8.98%	7.67%	7.67%	7.92%	8.21%	7.62%
2009	2.92%	2.02%	3.60%	1.72%	1.77%	2.45%	2.93%	1.04%
2010	1.47%	2.10%	2.06%	1.99%	2.07%	2.29%	2.35%	1.53%
2011	3.86%	3.41%	3.55%	3.47%	3.41%	3.39%	3.23%	3.85%
2012	5.81%	4.76%	5.33%	4.80%	4.56%	4.88%	4.95%	4.91%