# CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies


## Bachelor thesis


**2017** **Marek Šedivý**

# CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

**Marek Šedivý**


# New new bottom billion: Poverty and regional differences in price levels around the world

*Bachelor thesis*

Prague 2017

**Author**: Marek Šedivý

**Supervisor**: Petr Janský, PhD.

**Academic Year**: 2016/2017

## Bibliographic note

ŠEDIVÝ, Marek. *New new bottom billion: Poverty and regional differences in price levels around the world*. Prague, 2017. 62 s. Bachelor thesis (Bc.) Charles University in Prague, Faculty of Social Sciences. Institute of Economic Studies. Supervisor Petr Janský, PhD.

## Abstract

Comparable national price levels are a fundamental instrument for any research attempting to compare economic indicators of various countries. Nowadays they are produced by a number of organizations, namely the World Bank, Organization for Economic Cooperation and Development and the Penn World Tables. However these are provided only on national level even though significant evidence of inter-regional differences in price levels exists. This could lead to a bias of income-based economic indicators. Systematic assessment of the impact of spatial price differences on income-based economic indicators was done in this thesis. On the basis of the existing research it estimates regional price levels for a sample of 21 high and middle income countries. By combination of these indicators with income survey data provided by the Luxembourg Income Study it constructs Gini coefficients and poverty headcount ratios both adjusted and unadjusted for regional price levels. Significant and persistent evidence of bias induced by regional price levels is found throughout the sample. In the majority of cases the failure to adjust for spatial price differences leads to overestimation of income inequality and incidence of poverty. The overestimation is most significant in the case of middle income countries.

**Keywords:** Regional Price Levels, Poverty, Inequality, Purchasing Power Parity, Gini Coefficient, Poverty Headcount ratio

**Range of thesis:** 63 766 characters

# Abstrakt

Srovnatelné cenové hladiny jsou nezbytným nástrojem pro srovnání ekonomických indikátorů rozdílných států. V dnešní době jsou produkovány následujícími organizacemi – Světovou bankou, Organizací pro hospodářskou spolupráci a rozvoj a organizací Penn World Tables. Nicméně tyto cenové hladiny jsou poskytovány na národní úrovni a to vzdor tomu, že existence vnitrostátních rozdílů v cenových hladinách byla prokázána. Opominutí vnitrostátních rozdílů v cenových hladinách by mohlo mít za následek vychýlení odhadů ekonomických indikátorů založených na příjmech. Tato práce má za cíl odhadnout míru tohoto vychýlení. Za tímto účelem jsou využity již existující práce zabývající se touto problematikou a existující odhady regionálních cenových hladin. Tyto data jsou využity k sestavení ekonometrického modelu, s jehož pomocí jsou následně odhadnuty regionální cenové hladiny pro vybranou skupinu 21 států, jež jsou následně zkombinovány s daty pocházejícími z průzkumů příjmů poskytnutými organizací Luxembourg Income Study za účelem spočtení Gini koeficientů a procenta lidí žijících pod hranicí chudoby. Oba indikátory jsou spočteny upravené a neupravené o regionální cenové hladiny. Tímto postupem je odhaleno signifikantní a přetrvávající vychýlení těchto indikátorů způsobené regionálními cenovými hladinami. Pro většinu uvažovaných států vede zahrnutí cenových hladin ke snížení indikátorů chudoby a nerovnosti. Jejich nadhodnocení je největší u zemí, které jsou Světovou bankou klasifikovány mimo skupinu zemí s vysokým příjmem.



**Klíčová slova:** Regionální cenové hladiny, Parita kupní síly, Nerovnost, Chudoba, Cenové hladiny, Gini koeficient, Míra chudoby

**Počet znaků:** 63 766

## Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

2. The author hereby declares that all the sources and literature used have been properly cited.

3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague … **17. 5. 2017** **Marek, Šedivý**

## Acknowledgments

The author is grateful especially to Petr Janský, PhD. for his kind and patient supervision and valuable comments.

My gratitude also goes to prof. Zhong Zhao for the insight he provided into the Chinese labour market data, prof. Stef van Buuren for the consultation on the method of Multiple Imputation and to Bettina Aten for the comments and remarks she kindly provided.

I am also grateful to Luxembourg Income Study for providing me with access to their database.

Also I would like to acknowledge and thank for the help to all statistical offices that helped me during the construction of the datasets used in this thesis, namely – Central Statistical Office (Ireland), Statistics Denmark, National Statistics Office of Georgia, Turkish Statistical Institute, Philippine Statistics Authority and General Statistics Office of Vietnam.

**Institute of Economic Studies**

**Bachelor thesis propsal**

**Supervisor:** Petr Janský, PhD.

**Proposed topic**: New new bottom billion: Poverty and regional differences in price levels around the world

# Characteristic

***Research question and motivation***

Price levels have a significant impact on the standards of living among states or regions. Hence their knowledge is essential for the estimation of the distribution of world´s poverty. During the elaboration of such estimates national purchase power parities (PPPs) are used to convert data from national currencies to comparable units so that monetary statistics can be compared (e.g. GDP, income). PPPs serve both as convertors to common currency and they also equalize the purchasing power of different currencies. Nationwide PPPs are used even though there exist price level heterogeneities within states. Hence the usage of regional price levels would provide better estimates. Even though statistical offices of certain states have already collected data on regional price levels, these data are not collected worldwide. Therefore I will estimate them using an econometric model and then use them to estimate the distribution of world´s poverty. This estimation will be done for all the countries for which relevant data are available. With the distribution obtained this way I will see whether the estimation based on regional price levels differs significantly from the estimation based on nationwide price levels. Also I will observe if the knowledge of regional price levels and their application doesn´t change the eligibility for development aid of the receiving countries and finally based on the results I will draw conclusions for the development policies.

***Contribution***

The main contribution of my bachelor thesis shall be the estimation of distribution of world´s poverty based on regional price levels. The worldwide estimation of regional

price levels should also be of significance as to my best the knowledge the only time this exercise was done in Aten and Heston (2003).

***Methodology***

To estimate the regional price levels I will construct an econometric model with the price level of the given region as response variable. It will be an OLS model. For the construction of this model the identification of significant control variables will be crucial. In order to identify them, various econometric models will be constructed and the variables with highest significance chosen. Valuable insight to the identification of control variables and construction of the model will be provided by the works that have already estimated regional price levels e.g. Aten and Heston (2003), Čadil et al. (2014) and Roos (2006). Amongst the control variables that have been used in the previous studies to estimate regional levels have been – density of the population, income/annual wage, GDP per capita, tourism, education and unemployment. Hence these variables will be at the heart of the constructed model and a search for other significant variables will also be conducted. As the source of the data for control variables will serve mainly the statistical offices of the respective countries. Valuable data can also be provided by OECD, Eurostat, World Bank, United Nations Human Development Reports and by the Oxford Poverty and Human Development Initiative.

***Outline***

Introduction

1. Literature review
2. Variable definition and dataset description
3. Econometric model
4. Out-of-sample predictions
5. Distribution of income
6. Poverty measures

Conclusions

***Literature***

Aten, Bettina, and Alan Heston. 2003. “Regional Output Differences in International Perspective.” World Institute for Development economics research (UNU-WIDER), No. 2003/55.

Bajgar, Matěj, and Petr Janský. 2014. “Regionální Rozdíly v Kupní Síle: Ceny, Platy, Mzdy a Důchody.” Praha: Národohospodářský ústav AV ČR, v. v. i.,.

Beck, Günter W., Kirstin Hubrich, and Massimiliano Marcellino. 2009. “Regional Inflation Dynamics within and across Euro Area Countries and a Comparison with the United States.” Economic Policy, 24 (57), p. 142-184.

Čadil, Jan, Petr Mazouch, Petr Musil, and Jana Kramulová. 2014. “True Regional Purchasing Power: Evidence from the Czech Republic.” Post-Communist Economies, 26:2, 241-256. DOI: 10.1080/14631377.2014.904109.

Deaton, Angus, and Alessandro Tarozzi. 2000. “Prices and Poverty in India.” Research Program in Development Studies, Princeton University.

Edward, Peter, and Andy Sumner. 2015. “New Estimates of Global Poverty and Inequality How Much Difference Do Price Data Really Make.” Center for Global Development. http://www.cgdev.org/publication/new-estimates-global-poverty-and-inequality-how-much-difference-do-price-data-really.

Roos, Michael W. M. 2006. “Regional Price Levels in Germany.” Applied Economics. Vol. 38, No. 13, s. 1553-1565.

Sumner Andy. 2016. “Global Poverty and the New Bottom Billion Revisited Why Are Some People Poor.”

“UK Relative Regional Consumer Price Levels for Goods and Services for 2010.” 2011. Office for National Statistics. https://data.gov.uk/dataset/regional_consumer_price_levels.

# Contens

# Introduction

Price levels have a significant impact on the differences in standards of living among states or regions. That makes their knowledge essential for any attempts to estimate the distribution of income, indicators of inequality or poverty headcounts. In order to make monetary data stemming from different regions comparable Purchasing Power Parities (PPP) are usually applied. PPPs serve as convertors to common currency and equalizers of the purchasing power of different currencies. Nowadays three organizations produce PPPs – The Organization for Economic Cooperation and Development (OECD) in a joint project with Eurostat, International Comparison Project (ICP) of the World Bank (WB) and the Penn World Tables (PWT). Even though the provided price level measures are constructed by different methodologies they are all calculated on the national level. This is done even though substantial evidence of intra-national price level disparities exists. Compared to the attention given to intra-country temporal price differences, measured by the Consumer Price Index (CPI), spatial price differences gained little attention. To the best of my knowledge only two statistical offices in the world provide regularly measures of regional price levels – The Bureau of Economic Analysis (BEA) in the USA and the Turkish Statistical Institute (Turkstat). Therefore the solution proposed by a number of researchers and institutions is to estimate the regional price levels.

Current developments in the research into the distribution of income and global poverty made estimation and inclusion of regional price levels into economic research even more vital. Collier (2007) states that about a billion people lives in a set of 58 poorest countries. For these he coins the term *"The Bottom Billion"*. He argues these states fail to produce sufficient growth as a consequence of being locked in one or a combination of multiple of the following traps – conflict trap, natural resource trap, bad governance trap or being landlocked with bad neighbors. However Sumner (2011) showed what became a stylized fact nowadays - that about a billion or up to three quarters of global poor are located in middle income countries. Considering that states with large inter-regional differences such as China or India belong to this group, the omission to adjust for regional price levels could lead to significant bias. However statistics of high income countries are subject to regional price levels induced bias too. The impact of regional price levels in high income countries was shown for example by Aten and Figueroa

(2014) and Pittau et al. (2011) for the United States of America (USA) and Italy respectively. However poverty measures are not the only income-based indicators prone to bias induced by spatial price differences. Measures of income inequality could be biased too. The goal of this thesis is to measure the influence of regional price levels on the indicators of income inequality and poverty. For this purpose regional price levels of a selected sample of 21 high and middle income countries will be estimated and both regional price levels adjusted and unadjusted measures of poverty and income inequality will be produced.

The remainder of this thesis is organized as follows. The first part contains a brief literature review with an overview of different methodologies used for estimation of regional price levels. The literature review is followed by section two which contains the definitions of key terms and description of the data set used for construction of an econometric model explaining the determination of regional price levels. The model is constructed in section three. Its out-of-sample predictive powers are tested in section four. Section five presents the impact of regional price levels on measures of income distribution while section six discusses their impact on poverty measures. The final section concludes.

# 1. Literature review

The problematic of price levels was discussed thoroughly in academic literature. The discussion began with an inquiry into national price levels. However during the past two decades regional price levels received significant attention too. Four approaches towards their estimation were identified and applied. Therefore before the creation of the econometric model a literature review is provided to present the findings in the field of regional price levels.

The importance of price levels for economic research is well recognized. It has been the subject of many both theoretical and empirical studies. Probably the most notable contributions were made by the seminal papers of Balassa (1964) and Samuelson (1964). These established the well-known Balassa-Samuelson hypothesis. Their approach was based on the observation of persistence of higher price levels in countries with higher income. Possible explanation of this phenomenon may be the larger productivity in the sector of tradable goods. This relation has been a subject of numerous studies testing its empirical validity. Kravis and Lipsey (1988) use a dataset consisting of 60 countries, proving the relationship between productivity in the sector of tradable goods and the price level of the given country. Heston, Nuxoll and Summers (1994) adopt a larger dataset of 85 countries proving that the relationship holds. However regional income or GDP per capita were not identified as the only determinants of price levels. Ahec-Sonje and Nestic (2002) propose a set of influential variables based on a dataset of 39 countries from the 1993 European Comparison Program. These include the openness of the economy, dummy variable for transition economies and the size of the government sector. Gelb and Diofasi (2015) use a larger data set consisting of 168 countries from the 2011 round of the ICP. Openness of the economy and of the labour market, fuel subsidies, institutional quality and a set of geographic variables were identified as significant determinants of price levels. Existence of higher price levels in Sub-Saharan Africa states compared to equivalent economies was also found. Isis (2016) provides possible explanations including the lower statistical abilities of the given statistical offices which could have an impact on the quality of produced indices. Urban bias present in the underlying surveys and difficulties connected with inclusion of self-produced goods in them were also stated as possible explanations.

Despite the in-depth analysis received by the question of national price levels and their frequent application in economic research, much lower attention is paid to intra-country i.e. inter-regional differences in price levels. Theoretical explanation of existence of inter-regional differences in price levels is provided in Suedekum (2006). By adding a home goods-sector to the model outlined in the seminal paper of Krugman (1991) he shows that a core-periphery structure with higher price level in the core can appear. Even though the research into intra-country differences in price levels is not as extensive as in the case of national price levels, it has received significant attention and a number of studies tried to produce their estimates. A variety of approaches was adopted to do so.

The first approach toward estimation of regional price levels is the application of classical price level indices. Deaton and Heston (2010) provide a comprehensive description of the formulas and methodology used for production of PPP. Thorough discussion of the 2005 round of the ICP is also provided in their work. The OECD-Eurostat methodology of construction of PPPs was used in several works that tried to estimate the regional price levels. The Office for National Statistics (ONS) in the United Kingdom combined data collected for the construction of the consumer price index (CPI) with a survey on regional prices. The methodology identical to the OECD-Eurostat program was used and Relative Regional Consumer Price Levels (RCPL) were constructed. The RCPL for the UKs NUTS 1 regions are reported in *"UK Relative Regional Consumer Price Levels for Goods and Services for 2010" (2011)*. These show that the price level in the region of London is more than 10% higher than in the region of Yorkshire and the Humberside. The OECD methodology was also adopted by the Turkish statistical institute (Turkstat), which produces the regional price levels for the Turkish NUTS 2 regions. However this methodology was also applied in academic research.

Significant number of academic works using the above outlined methodology can be found. The Èltetö-Köves-Szulc method, applied by the joint project of OECD and Eurostat also inspired the approach used by Kramulová et al. (2016) who provide regional price levels for the 14 Czech NUTS 3 regions based on the data collected for the construction of the CPI. Significant difference of 25,6% in the price levels of the region Prague and the Ústecký region are reported. Deaton (2003) uses the consumption

data from the 43rd, 50th and 55th round of the Indian National Sample Survey (NSS) to construct Laspeyers, Paasche, Fischer Ideal and Törnqvist price indexes for the rural and urban parts of the 17 Indian regions. These price levels are then used to show that the reduction in poverty headcount that occurred in between the different rounds of the NSS was partly caused by the underlying changes in regional price levels. A different approach was adopted by Brandt and Holz (2006) who profited from the data published in the Price Statistical Yearbook of the Chinese National Bureau of Statistics (NBS) to estimate the price levels for the rural and urban parts of the 16 Chinese regions as well as one combined price level for each region. It is shown that the adjustment for the price levels leads to a decrease of 30% in the 1990 Gini coefficient. Li, Zhang, and Du (2005) also provide regional price levels for the regions of China showing significant inter-regional differences. According to their data, in 2002 the price level in the province of Qinghai was only 61% of the price level of the Shanghai province. Radvansky and Fuchs (2012) provide regional price levels for the Slovak NUTS 3 regions. Since 2008 the Bureau of Economic Analysis (BEA) in a joint project with the Bureau of Labour Statistics (BLS) constructs comparable regional price levels for the 50 states of the USA and the District of Columbia. Aten (2015) describes in detail the methodology through which data collected for the construction of regional CPI are transformed into comparable price levels. Based on the data used for construction of the CPI, price levels of the Philippines regions were estimated during a one-off exercise conducted by the Asian Development Bank (ADB). Dikhanov, Palanyandy, and Capilit (2011) report the resulting regional purchasing power parities for the years 2005 through 2010.

Another frequently used approach is hedonic regression, a variation of the Country Product Dummy method proposed by Summers (1973). Aten and de Menezenes (2002) used it in their pioneering work to estimate regional price levels for eleven cities of Brazil. For their estimation they have used the data from household expenditure survey. Hedonic regression was also used by Aten and Heston (2003) in what remains until now the largest estimation of regional price levels. New approach was created by Coondoo, Majumder, and Chattopadhyay (2011) who estimated regional price deflators for the rural and urban parts of the 17 Indian regions by creating a demand system based on the NSS dataset and a subsequent analysis of the Engel curve. This approach became popular among the researchers estimating regional price levels for the Indian regions. Most recent applications include for example Majumder and Ray (2015) or Majumder,

Chakrabarty, and Ray (2014) who use more actual data from the NSS. Engel curve analysis approach was also adopted by Majumder, Ray, and Sinha (2011) who provide estimates of regional price levels for the regions of India and Vietnam. Mishra and Ray (2014) use the *"Exact Affine Stone Index"* to estimate regional price levels of the 7 regions of Australia. Significant difference of more than 90% is found in between the price levels of the regions of Western Australia and Tasmania.

Roos (2006) adopts an approach different from the two outlined above. He uses the dataset provided by Ströhl (1994) to construct an OLS model with the price levels as controlled variable. He then uses it in second step to perform out-of-sample predictions. Filling in the missing values this way he estimated the regional price levels for the 16 German NUTS 1 regions for the year 1993. These are then shifted with the use of regional CPI indices to 2002 showing for example a 7% difference in between the price level of Berlin and the region of Brandenburg. Identical approach was adopted by Janský and Kolcunová (2017) for the NUTS 2 regions of EU states. Based on the estimated price levels the authors report possible misclassification of the regions in the framework of the EU cohesion policy. However Blien et al. (2009) criticized this approach asserting that the construction of the econometric model based on estimated values for which standard errors are unknown leads to a bias in the estimated standard errors. As a solution the authors propose a forth possible method for estimation of regional price levels, application of the Multiple Imputation.

Once estimated, the regional price levels have a significant impact on many economic statistics. The most frequent application is the comparison of regionally adjusted income differentials. Aten and Figueroa (2014) report the impact of adjustment for regional price differences on income distribution for the states of the USA. Bajgar and Janský (2014) use the set of regional price levels provided by Čadil et al. (2014) to show the differences between nominal and real regional income for the regions of Czech Republic. Pittau, Zelli, and Massari (2011) use the regional deflators constructed by the Italian Statistical Office and the Union of Italian Chambers of Commerce for the same objective for the regions of Italy. Blien et al. (2009) test whether the rural urban wage differential persists even after the differences in regional price levels are accounted for, proving that even though there is a stable differential it is not as high as it may appear from data unadjusted for regional price levels.

# 2. Variable definition and dataset description

Definition of key terms is essential for the construction of an econometric model. Therefore this section provides the definition of fundamental terms. Both dependant and independent variables are presented and the sources of data outlined. The issues encountered during the construction of the dataset are also presented.

## 2.1 Dependent variable

Price level is an economic variable that is supposed to express the price level of one country relative to another. At the national level it is usually constructed by division of the PPP by the nominal exchange rate. PPP serve as indicators of comparable cost among different countries. They serve as both convertors to same currency and spatial price deflators. While referring to comparable spatial price levels the term regional price level will be used in this thesis. It can also be interchanged for the term regional purchasing power parity.

To the best of my knowledge, only three statistical offices in the world have ever provided regional price levels, the BEA in cooperation with the BLS for USA, the Turkstat for Turkey and the ONS for the UK. Therefore sources from academic literature had to be adopted. Regional price levels for fourteen countries, including the above described, were found. These were not constructed by the same methodology and also vary significantly in the size of the regions concerned. The fact that certain regions within the sample are of the size of whole countries, calls for regional adjustment which will be described latter on.

Even though actual or estimated regional price levels exist for fourteen states, only twelve of those estimates could have been used during the construction of the dataset. Unavailability of control variables on the appropriate regional level made the estimated price levels for the regions of Vietnam provided by Majumder, Ray, and Sinha (2011) unusable. Non-existence of the control variables for the regions was confirmed also by the General Statistics Office of Vietnam. Even though a significant number of estimates exist for the regions of India these are all based on the National Sample Survey. Therefore they provide regional price levels for rural and urban parts of the selected regions separately. However only the variables related to labour market are provided by

the Indian Ministry of Statistics and Program Implementation on this level. Possible solution would be a creation of an artificial price index representing both the rural and urban sections of a given region. This solution was rejected as it would artificially introduce more imprecisions into the dataset. Therefore the available regional price levels for Indian regions were not used either.

As sources from academic literature were adopted the question of selection among various data sources had to be answered. The rule of thumb approach that was adopted was to prefer the more recent estimates in place of older ones. Another important criterion was the quality of underlying data. Choice between various estimates had to be done for the states of China and Germany. In the case of China the choice was done between Li, Zhang, and Du (2005) and Brandt and Holz (2006). Even though the former provides regional price levels constructed on a larger and more actual basket than the latter, the latter was chosen instead. The underlying reason is that Li, Zhang, and Du (2005) provide regional price levels respective to Shanghai instead of the whole country. After consultation with the authors I was assured that transformation of the data into the desired form was not possible. As can be seen from Figure 1 estimates provided by the latter study significantly underestimate regional price levels at least in seven regions. This issue is treated during the construction of the model in the following section. In case of Germany two competing estimates exist, namely Roos (2006) and Deckers et al. (2013). Even though the price levels constructed by the latter study are based on a more recent dataset, the price levels provided by the former one were chosen. This choice was caused by the fact that the latter dataset provides regional price indices over a period of time instead of indices constructed for a particular year as is the case in other works used in the dataset.

Even though different methodologies have been employed during the construction of the regional price levels, all the estimates can be considered as sufficiently representative of the differences in between the regional price levels. As a result the below outlined model is based on 227 regional price levels. The complete list of sources of the regional price levels with description is provided in Table 1.

## 2.2 Explanatory variables

During the construction of the econometric model assumptions identical to Roos (2006) and Janský and Kolcunová (2017) were adopted. Under the following assumptions the regional price levels are determined only by the differences in regional supply and demand. These assumptions include the spatial segmentation of regional markets which makes impossible any strategic price setting or arbitrage. Also consumers and firms are considered to be immobile in the short run. Furthermore intermediate inputs are traded among the regions at no transportation cost and have the same price in each of the regions.

The selection of explanatory variables was inspired by the econometric models constructed in Roos (2006), Janský and Kolcunová (2017) and Blien et al. (2009). Valuable insight into the selection of control variables was provided also by Gelb and Diofasi (2015). Regional disposable income may be considered as the major determinant of the strength of regional demand. Furthermore the number of consumers living within the given region shall be highly correlated with the strength of regional demand. Population density may be of influence for example through underlying quality differences, associated mainly with the service sector. Area and GDP per capita are also tested as explanatory variables. Set of variables describing the labour market is also considered. These include the employment, unemployment and participation rate. A set of dummy variables describing the characteristics of the region is also proposed. Dummy variables indicating the presence of the capital city within the region or a city above 1% or 2% of total populations are also tested. Access to the sea or a presence of international airport in the region could translate into a higher amount of visiting tourists, which can be considered as an important source of demand. Dummy variable indicating the presence of a monument from the UNESCO world heritage list is also proposed to model the influence of tourists on the regional price levels. Dummy variable for Chinese regions was also included to control for the above described imprecisions in the dependant variable as well as imprecisions in the control variables that are described below. Full list of proposed control variables is included in Table 2.

Gelb and Diofasi (2015) suggest that price levels could also be determined by institutional quality. Unfortunately no usable measure of intra-country differentials in

institutional quality was found. Even though a possible solution would have been to include an indicator of national level institutional quality, this solution would require transformation of the data. Instead of defining the regional price level as the price level of the region respective to the national one, an alternative definition would have to be considered. Given that institutional quality would be modelled at national level, the explanatory variables and dependent variable would have to be transformed to measure the differences compared to a global indicator e.g. average of all regional values. However considering that the purpose of this thesis is the calculation of national economic indicators this approach was rejected. Therefore the influence of institutional quality on the regional price levels was not tested.

### 2.2.1 *Regional adjustment*

The fact that certain regions in the dataset are of the size of some included states combined with estimation of a relative measure such as regional price level calls for a transformation of explanatory variables. As a consequence the variables were transformed according to the formula proposed in Janský and Kolcunová (2017) –

$$x_i = \frac{regional\ value}{average\ national\ value} * 100$$

### 2.2.2 *Data issues*

Despite the fact that the majority of the data needed for the construction of the dataset is available from public sources or upon request from the statistical offices of the respective states, some problems were encountered. Philippine Statistics Authority (PSA) publishes annually the Philippine Statistical Yearbook (PSY) containing the summary statistics about the state of Philippines and its regions. However some time series are provided only in three year intervals thus making the construction of the dataset for one base year impossible due to unavailability of data. The majority of data is available for the year 2010 except disposable income, which has been released for the years 2009 and 2012. Filling in the data set by data constructed with the following formula was considered.

$$inc_{2010} = inc_{2009} + \frac{1}{3} * (inc_{2012} - inc_{2009})$$

However as can be seen from Figure 2 there have been minimal changes in the regional differences in the distribution of disposable income between the years 2009 and 2012.Therefore during the construction of the model outlined below, the data for the year 2012 were used as these are deemed to represent well enough the differences in the disposable income between regions.

Labour market statistics for Chinese regions represented another issue that had to be addressed during the construction of the dataset. The Chinese National Bureau of Statistics (NBS) provides only data indicating the regional unemployment rate. However sufficient data are provided in the Statistical yearbook published by the NBS to construct both participation and employment rates. Even though these were constructed an important caveat remains. Labour market statistics and mainly the unemployment rate were judged as highly unrepresentative by a number of academic studies such as Giles et al. (2005) or Lian (2012). Cai et al. (2013) even marked them *"almost useless"* as they are likely to significantly underestimate the true values. Feng et al. (2015) identify the following reasons for the underestimation. First a large fraction of the people lacks household registration, the so-called *"hukou"*, which makes them unqualified to register with local employment service agencies. However even qualified people may not register with the authorities due to low levels of unemployment benefits. Furthermore the aggregated data could be subject to aggregation errors due bottom-up aggregation of registered unemployed people. For these reasons a search for an alternative source of labour market indicators was conducted. Even though estimation of national unemployment rates for China on the national level was done no source of regional unemployment rates for China was found in academic literature. Therefore the labour market indicators based on the Statistical Yearbook published by the NBS were used. As was already noted above, the imprecisions in the Chinese data are controlled for by a dummy variable for Chinese regions.

# 3. Econometric model

Variety of approaches towards the estimation of regional purchasing power parities could have been adopted. These have been outlined in the literature review and include CPD methodology, Multiple Imputation, construction of an OLS model which would be used for out-of-sample predictions and production of spatial indices based on data collected for production of regional CPI. The CPD methodology had to be abandoned due to its high demand on input data. Construction of spatial indices based on data collected for creation of temporal ones was rejected for the same reason. Therefore two feasible approaches were left, namely the Multiple Imputation framework and construction of an OLS model. The former framework was rejected for the following reason. Blien et al. (2009) used it for filling in of missing variables in a dataset which they used in a second step to perform linear regression. For such applications this approach guarantees convergence of the estimated coefficients to their true value. However the purpose of this thesis is to use the regional price levels for calculation of adjusted indicators, not for further econometric estimation. According to van Buuren (2012) in such applications the Multiple Imputation produces standard errors that are too short. Considering that this was the main reasons for which Blien et al. (2009) preferred Multiple Imputation instead of out-of-sample prediction by an econometric model the OLS framework was adopted. However this approach has certain drawbacks. Application of a general equilibrium model might be more appropriate as some of the explanatory variables may be determined simultaneously. Instrumentation for the endogenous variables could be a possible solution too. However employment of these methods was prohibited by the available data. Another possible approach towards the construction of the model would be to create a single time series model for each country. As it was already noted in the previous section, the underlying dataset consists of 227 regional price levels observed in 12 states. Even though this dataset is considered to be sufficiently representative of the mechanisms underlying the determination of regional price levels, panel data are available only for the USA. Under the assumption that the effect of the explanatory variables is the same in all countries a linear model was constructed by OLS regression.

After running a series of regressions it was found that a level-level model provides the best fit for the data. Following algorithm of model construction was adopted – all

variables outlined in Table 2 were regressed on the dependent variable and the one with highest statistical significance was kept in the model. The remaining variables were then added successively to the already chosen variable in order to identify the most significant one, which would be added to the model. These steps were repeated until none of the remaining variables was found to be of statistical significance. Model 1 in Table 3 is the constructed model. Following variables were found to be statistically significant – disposable income, percentage of total area covered by the region, population density and unemployment rate. Dummy variable for Chinese regions introduced because of imprecisions in Chinese data also tested as highly significant. As can be seen only 40% of the variation in the dependent variable is explained by the fitted model. Even though this may seem low compared to previous applications of this methodology, the significant heterogeneity of the considered sample should be borne in mind. The fit of the resulting model cannot be as high as in the case of a more homogenous sample of countries such as member states of the EU or regions within a particular state. To discuss the quality of the model the assumptions necessary for its construction had to be tested.

Gauss-Markov assumptions were tested in order to assess the quality of the model outlined in the previous paragraph. The possibility of functional form misspecification was addressed as first. Ramsey regression specification error test (RESET) test was conducted. Enough evidence for possible misspecification was not found even at the 40% significance level. Possible presence of heteroskedasticity remains the last thing to test for as endogeneity was addressed already in the first paragraph of this section. Figure 3 presents the summary plots of Model 1. From the plot of standardized residuals we may see that heteroskedasticity could be present. However a more rigorous test is needed. White´s test for heteroscedasticity was conducted and the null hypothesis of homoscedasticity was rejected at 1% significance level. Therefore it can be concluded that there is enough evidence for presence of heteroskedasticity in the model. Even though heteroskedasticity does not cause the OLS estimator to be biased it has a significant influence on hypothesis testing. Furthermore under the presence of heteroskedasticity the OLS estimator is no longer the Best Linear Unbiased Estimator (BLUE). Even though sufficient evidence was not found to prove violation of certain assumptions, the presence of heteroskedasticity has to be treated.

Table 3: Estimation results

| Variable | Model 1 | Model 1 (heteroskedasticity robust) | Model 2 |
|---|---|---|---|
| **Intercept** | 91.0053157 *** (2.6245335) | 91.00531573 *** (3.73929042) | 94.11435666 *** (2.28025817) |
| **Increl_2** | 0.0007545 *** (0.0001211) | 0.00075448 *** (0.00012413) | 0.00071186 *** (0.00011342) |
| **UnempRrel** | -0.0353963 . (0.0188330) | -0.03539633 * (0.01709411) | -0.04399674 ** (0.01489588) |
| **Popdensrel** | 0.0076046 ** (0.0029061) | 0.00760455 (0.00506371) | 0.00410546 . (0.00213624) |
| **Areaperc** | 0.6163452 *** (0.1137674) | 0.61634523 (0.43811017) | |
| **Popperc** | | | 0.35627067 ** (0.13281582) |
| **China** | 11.5033999 *** (1.6716617) | 11.50339993 * (2.01340691) | 10.94029673 *** (2.09056464) |
| **Adjusted R-squared** | 0.4038 | 0.4038 | 0.3494 |
| **F-statistics** | 31.61 | 31.61 | 25.27 |

"***" $p < 0{,}001$; "**" $p < 0{,}01$; "*" $p < 0{,}05$; "." $p < 0{,}1$

In the presence of heteroskedasticity the classical OLS inference is not possible. Thus it has to be treated. The main consequence of heteroskedasticity is that t-statistics do not have t-distribution and F-statistics do not have F-distribution. As a consequence testing of statistical hypothesis is no longer possible. Presence of heteroskedasticity can be treated by a set of instruments. One possibility would be to apply Weighted Least Squares regression. However this solution is not feasible as the form of heteroskedasticity is unknown. Feasible Generalized Least Squares regression could be adopted too, but it would produce biased estimates. White´s standard errors are therefore used. Model 1 is re-estimated and White´s standard errors are calculated in

order to test for the significance of the parameters. The results are reported in Table 3. Once heteroskedasticity is controlled for a significant drop of statistical significance of population density and the percentage of area covered by the given region occurs. These are no longer significant even at the 10% level. New model was also constructed by the same algorithm as Model 1 but with the use of White´s standard errors. The resulting model is also reported in Table 3 as Model 2. The usage of White´s standard errors caused the variable indicating the percentage of area covered by the region to become insignificant. Percentage of population living in the region became significant instead. Only nearly 36% of the variation in the dependent variable is explained by Model 2. This is at least partly caused by the numerous outliers present in the data set. Analysis of their effect on the final estimate is thus necessary.

Significant number of outliers can be identified within the dataset. This can be seen also from Figure 4 which shows the boxplot of regional purchasing power parities. Cook´s distance was used for identification of outliers and high leverage observations. All observations with the Cook´s distance higher than the conventional cut off point 4/n, where n is the number of observations in the dataset, were considered as outliers or high leverage observations. Thirteen outliers were identified, namely the regions of Prague (CZ), Scotland (GB), Zhejiang (CN), Hainan (CN), Chongqing (CN), Yunnan (CN), Qinghai (CN), Manila (PH), Autonomous region in Muslim Mindanao (PH), Queensland (AU), Northern Territory (AU), Western Australia (AU) and the District of Columbia (US). To see whether the outlying observation had a significant impact on the resulting model a new model was constructed based on the data set without outliers. It is reported in Table 4 as Model 3. However even in this model heteroskedasticity was detected and White´s standard errors were used. Heteroskedasticity robust results are also reported in Table 4. By the comparison of Model 3 with Models 1 and 2 it is obvious that the outlying observations have a significant impact on both statistical and economical significances of control variables. Furthermore the goodness of fit is highly affected too. Robust regression methods were used to account for the influence of outliers too. Huber weighting function was used and a robust liner model constructed. The results of the robust regression are reported in Table 4 as Model 4. High influence of outlying observations is also indicated by the results of the robust linear model. Their high influence has to be considered during the selection of the model used for out-of-sample predictions.

Table 4: Outlier analysis

| Variable | Model 3 | Model 3 (heteroskedasticity robust) | Model 4 |
|---|---|---|---|
| **Intercept** | 91.86 *** (1.809) | 91.862 *** (1.9597) | 92.3085 *** (1.62) |
| **Increl_2** | 0.0007727 *** (0.00008368) | 0.00077266 *** (0.000095907) | 0.0008 *** (0.0001) |
| **UnempRrel** | - 0.03126 * (0.01273) | -0.031264 * (0.013515) | -0.033* (0.0128) |
| **Popdensrel** | 0.01195 *** (0.003189) | 0.011952 ** (0.004391) | 0.0062 ** (0.002) |
| **Areaperc** | 0.2309 * (0.09015) | 0.23090** (0.088444) | 0.2441 ** (0.0774) |
| **China** | 9.539 *** (1.153) | 9.539 *** (1.7705) | 10.18 *** (1.1377) |
| **Adjusted R-squared** | 0.5821 | 0.5821 | |
| **F-statistics** | 60.34 | 60.34 | |

“***” p < 0,001; “**” p < 0,01; “*” p < 0,05; “.” p < 0,1

Both the influence of outliers and presence of heteroskedasticity ought to be considered during the selection of the model used for out-of-sample predictions. As was shown in the previous paragraph, outliers and high leverage points were of high influence on the resulting estimates. Furthermore the presence of heteroskedasticity had to be controlled for by the introduction of White´s standard errors. Considering both questions, Model 3 was judged to be the best for out-of-sample predictions as it does not suffer from bias caused by outliers and heteroskedasticity is controlled for too.

# 4. Out-of-sample predictions

The econometric model in the previous section was constructed for the purpose of out-of-sample prediction. Price levels constructed by it will be used for construction of both regional price levels adjusted and unadjusted inequality and poverty measures. Obviously if statistical offices provided measures of spatial price differences there would be no need for this approach. However, to the best of my knowledge these are provided regularly only by two statistical offices in the world. Thus the only feasible solution is estimation. Nevertheless before performing the out-of-sample prediction the predictive powers of the selected model should be tested.

## 4.1 Test of predictive powers

Various techniques were used for the test of the predictive powers of model 3. However before any calculations were performed the fitted values were plotted against the actual ones. The resulting plot is shown in Figure 5. It can be seen that even though the model does not precisely predict the actual values, it tends to simulate the differences in the regional price levels quite well. The fitted values obtained during the construction of the model were also used for calculation of the Root Mean Square Error (RMSE) which was found to be 5,32. Nonetheless as the RMSE tends to overvalue outlying observations given the heterogeneity of the considered sample of countries and presence of outliers in the data set another measure evaluating the goodness-of-fit was also used, namely the Mean Absolute Error (MAE). It was calculated and found to be 4,01. Even though lower values of these measures indicating better fit to the data would be desired, taking into account the heterogeneity of the states included in the dataset, these were judged as sufficient. However the fit to the data is not the main criterion to be considered as rather its out-of-sample predictive powers were the main motive for its construction.

As the main purpose of the construction of the Model 3 was to use it to perform out-of-sample predictions the fit to the dataset is not a sufficient criterion to judge its quality. Therefore a test of the predictive powers had to be conducted. The following approach was adopted – one by one, each of the observations present in the dataset was excluded and the model constructed each of the restricted samples. The omitted observation was then predicted with the model. 214 predicted values were obtained by this approach and RMSE was calculated. It was found to be 5,52. By comparison with the RMSE of the

regression based on the full data set we can see that these measures are not highly affected by exclusion of data points. Therefore the Model 3 will be used and the regional price level of region *i* will be predicted by the following equation:

$$PPP_i = 0.00077266 * income_i^2 - 0.031264 * unemployment_i + 0.011952 * population\ density_i + 0.2309 * area_i + 9.539 * China_i + 91{,}862$$

## *4.2 Comparison with existing estimates*

Representativity of the predicted regional price levels should be evaluated before their further application. For that purpose they were compared with the official indices provided by the ONS for the UK and the BEA for the USA as these are the only states whose statistical offices constructed regional price levels for the selected year. In the case of the UK the model had an overall tendency to overestimate the regional price levels. The average bias was 0,31% with North East England, East Midlands and Northern Ireland suffering from the biggest underestimations whereas Scotland and London suffered from the biggest overestimations. Figure 6 shows both the official and estimated indices for the regions of UK. Official and predicted regional price levels for the states of USA are presented in Figure 7. The average bias was -0,47%. The biggest overestimations of the regional price levels occurred in the cases of the District of Columbia and South and North Dakota. This can be explained by the fact that no variable indicating the differences in rents was included in the model as according to Aten (2015) the high price levels of District of Columbia and Hawaii are due to high rents. The states of California, Wyoming and New York suffered from the biggest underestimation. Even though the regional price levels for the UK and USA were calculated by the constructed model both the predicted and actual values will be used for further calculations. Indicators based on the latter will be discussed primarily as the official indices should be more representative of the differences in regional price levels.

Comparison of the constructed regional price levels with the estimates provided by other academic literature was also conducted. Janský and Kolcunová (2017) provide regional price levels for the states of European Union. Regional price levels on the same regional level were calculated for nine of these states in this thesis. By comparison of the results it is visible that the regional price levels estimated in this thesis tend to be of

lower value. However the absolute differences in between the regional price levels tend to be of the same magnitude. Therefore the estimated regional price levels may be considered as sufficiently representative of the spatial price differences and will be used for the assessment of spatial price differences induces bias in the measures of inequality and poverty.

Even though the chosen model was found to be sufficient for prediction of regional price levels it could be improved. The main weakness is exclusion of any variable indicating the housing prices within a given region as these were found to be significant determinants of the regional price levels. However due to unavailability of a unified measure for all states in the dataset, they could not have been included in the model. Unavailability of a unified indicator of the volume of tourists visiting the region also restricts its predictive powers as tourists serve as an important source of demand. Dummy variable indicating the presence of a monument from the UNESCO world heritage list was included to model the regions attractiveness to tourists. However it was unlikely to sufficiently model the differences in between the regions. Despite these weaknesses the model is still judged as sufficiently strong and was used for out-of-sample prediction of regional price levels.

# 5. Distribution of income

The crucial role played by purchasing power parities in economic research was already recognised by a significant number of researchers. Even though national price levels are frequently used in both academic and non-academic literature, their subnational counterparts received significantly lower attention. Regional price levels are not applied even despite various studies such as Brandt and Holz (2006) proved their impact on measures of income distribution. However this study provides a proof of their impact only for the China. To the best of my knowledge an assessment of impact of the regional price levels on the measures of income inequality for multiple states was not conducted yet. In the following section regional price levels of a selected sample of 21 countries are estimated and their impact on a measure of income inequality, namely the Gini coefficient is evaluated though first a brief description of the included countries and data sources is provided.

## 5.1 Luxembourg Income Study database

Construction of robust indicators of economic inequality requires reliable and representative microdata. The income surveys gathered and harmonized by the Luxembourg Income Study (LIS) were used. LIS database provides datasets for 49 countries and even though a variable indicating the region of origin of the given data point is included in the list of variables, it is not provided in each dataset. Another data restriction was imposed by unavailability of control data necessary for estimation of regional price levels for all countries in the LIS datasets with the regional variable filled in. Furthermore the regional disaggregation of the data is not provided on the same level for all states. So despite the fact that Eurostat NUTS 2 levels or OECD territorial level 2 would be preferred, the sizes of regions vary. For example the regional information for France is provided on NUTS 1 level whereas the information for Czech Republic or Slovakia is provided on NUTS 3 level. Complete list of states with sufficient information for both calculation of regional price levels and assessment of their impact on inequality measure is provided in Table 5. Both the regional price levels predicted with the use of Model 3 as well as the number of observations within each region are included. The sources of data used for prediction are outlined in Table 6. Few further adjustments were made to prepare the data for calculation of indicators of income inequality. All negative values present within the dataset were set to zero. These are

usually caused by self-employment. No other top or bottom coding was applied. Also only observations for which the region was indicated were kept in the data. This led to exclusion of 193 observations from the data set of Canada but even despite this, sufficient number of observations remained for each region for the sample to be representative. As can be seen the samples upon which the Gini coefficients are constructed remain sufficiently large and therefore can be considered as representative. The year 2010 was chosen for the majority of countries because of the availability of control variables.

## *5.2 Gini coefficient*

Despite the existence of superior measures of income inequality the Gini coefficient was used in this thesis. The main reason for its selection is that it is one of the most frequently used indicators of income inequality. Therefore the influence of regional price levels on it should be estimated. The Gini coefficient can be expressed graphically as the area between the Lorenz curve depicting the distribution of income within a society and the 45 degree line indicating perfect equality. Thus Gini coefficient equal to zero indicates perfectly equal distribution of income whereas Gini coefficient equal to one indicates a perfectly unequal society. Variety of income measures can be used for its construction. Three of these measures have been used in this thesis. Household disposable income was the first one. However Gini coefficients based on it may not sufficiently represent the distribution of income in society as the size of the household matters too. For this reason the disposable income per capita was also used. Even though Gini coefficients based on it take into account the size of the given household an argument for a better indicator can be made as the per capita income fails to account for the possible occurrence of economies of scale within households. Therefore the equalised disposable income was also used. It is obtained by division of the household disposable income by the square root of household members. Thus the disposable income diminishes more slowly with each new member of a household. This measure of income is proposed and commonly used by both the OECD and LIS as the basis for construction of Gini coefficients. Therefore most attention will be given to evaluation of the impact of the regional price levels on it but an assessment of impact of the regional price levels on the disposable household income and income per capita Gini coefficients is of value too.

## *5.3 Impact of regional price levels on Gini coefficients*

Both regional price levels adjusted and unadjusted Gini coefficients were calculated for all the countries in the sample. The results with precision to three decimal points are reported in Table 7. Considering the Gini coefficients based on equalised household income the least unequal states are Denmark, Slovakia and Czech Republic. Compared to that, Colombia, India, Mexico and Georgia were identified as the states with biggest income inequality. This ranking is not influenced by the adjustment for regional price levels. Overall the results indicate that regional price levels could possibly cause an upward bias of the Gini coefficients. From all the measures in 79,4% of the cases the adjustment for regional price levels led to lower estimates of inequality within the given country. Only for 14,3% of the indicators their application lead to no adjustment and in 6,3% of the cases the unadjusted Gini coefficients underestimated the income inequality in the state. Even though these numbers indicate a possible overestimation of inequality caused by regional price levels a more precise discussion of the results by type of income indicator will be done. For 71,4% of Gini coefficients based on household disposable income the application of regional price levels led to their decrease whereas for 28,6% of them the adjustment resulted in no change. No Gini coefficients based on the household income were underestimated. Gini coefficients based on per capita disposable income were adjusted downwards in 76,2% of cases though upwards adjustments were done only for 9,5% of the sample and the measure remained unchanged for 14,3% of the involved states. The same amount of states was adjusted upwards in the case of the Gini coefficients based on equalised income. However more indicators i.e. 14,2% remained unchanged. For the rest of the sample the application of regional price levels on the Gini coefficients based on equalised household income led to lower estimates of income inequality. Table 8 presents both the adjusted and unadjusted Gini coefficients based on the equalised income. The average value of adjustment was -0,0031, -0,0043 and -0,0039 for the disposable household, per capita and equalised income based Gini coefficients respectively. Therefore it may be concluded that at least in the restricted sample of countries upon which this thesis was based the regional price levels tend to cause an upward bias of the Gini coefficients. Figures 8, 9 and 10 represent the impact of regional purchasing power parities on the different indicators of inequality. Special cases within the data should be commented as well as discussion by country groupings ought to be provided. The only countries for

Table 8: Equalised income based Gini coefficients

| State | Adjusted | Unadjusted |
|---|---|---|
| **Australia** | 0,332 | 0,334 |
| **Austria** | 0,28 | 0,28 |
| **Canada** | 0,301 | 0,302 |
| **Colombia** | 0,496 | 0,511 |
| **Czech Republic** | 0,251 | 0,256 |
| **Denmark** | 0,255 | 0,254 |
| **France** | 0,292 | 0,292 |
| **Georgia** | 0,427 | 0,446 |
| **Germany** | 0,286 | 0,288 |
| **Greece** | 0,327 | 0,328 |
| **Hungary** | 0,281 | 0,293 |
| **India** | 0,499 | 0,503 |
| **Ireland** | 0,298 | 0,296 |
| **Italy** | 0,323 | 0,331 |
| **Mexico** | 0,459 | 0,465 |
| **Poland** | 0,312 | 0,314 |
| **Slovakia** | 0,262 | 0,264 |
| **Spain** | 0,335 | 0,337 |
| **Switzerland** | 0,299 | 0,299 |
| **UK** | 0,335 | 0,339 |
| *based on ONS RCPL* | 0,336 | |
| **USA** | 0,372 | 0,373 |
| *based on BEA RPP* | 0,372 | |

which the adjustment for regional price levels led to none or upward adjustments of the indicators of income inequality were Ireland and Spain. Largest adjustments had to be done to the following states – Georgia, Hungary, Mexico and Columbia. Considering that these are states with large inter-regional differences this results tend to support the hypothesis that significant differences in regional price levels lead to bias in Gini coefficients. However it is important to note that the size of the adjustments might be caused by the regional level on which the adjustment was done. This was NUTS 3 for

Hungary and OECD territorial level 2 for Columbia and Mexico. Surprisingly the adjustments done to the states with largest spatial price differences in absolute terms, namely India, France and USA were not as significant. The effect of adjustment for regional price levels on the measures of inequality should also be discussed by the type of states involved. For that purpose the WB classification was adopted. The sample consists of seventeen high income countries – Australia, Austria, Canada, Czech Republic, Denmark, Germany, Greece, Hungary, France, Italy, Ireland, Poland, Slovakia, Spain, Switzerland, UK, USA and four middle income countries – India, Mexico, Georgia and Columbia. The average adjustment for the equalised disposable income based Gini of the middle income countries was -0,00231 compared to only -0,011 for the high income states. Differences in the scale of adjustments are of the same magnitude for the remaining measures too. Despite the small sample of countries this implies possible overestimation of income inequality within middle income countries. The overestimation is more likely in these states as given their size and the disparities within their regions the differences in price levels are of higher magnitudes. However it is important to note that the size of the state is unlikely to be the principal determinant of the scale of adjustments to the Gini coefficients as for example in the case of other large states such as Germany or the USA the impact of regional price levels was significantly lower.

By calculation of regional price level adjusted and unadjusted Gini coefficients for 21 states it was shown that Gini coefficients can be subject to bias induced by regional price levels. This bias is the largest for states with significant inter-regional disparities. Countries not classified as high income states are also more prone to imprecisions caused by spatial price differences. Therefore adjustment for disparities in regional price levels should be done during the calculation of Gini coefficients. Given that Gini coefficients are subject to bias caused by regional price levels an inquiry into their influence on other statistics should be done too.

# 6. Poverty measures

Impacts of regional price levels on the measures of income distribution were already shown and discussed in the previous section. However these may not be the only statistics influenced by spatial price differences. Another possible consequence of omission to adjust for their differences could be a bias in the estimates of poverty headcounts as they may lead to overvaluation of income in regions with high price levels and undervaluation of income in low income regions. Hence the impact of adjustment for regional price levels on measures of poverty was tested. Definition of an appropriate poverty threshold was necessary for realization of this exercise.

## 6.1 Considered poverty lines

Term poverty can be used to accurately describe a number of situations varying from material deprivation to lack of opportunities. Given the nature of this thesis only income poverty is considered. Therefore a person will be identified as poor if her income is below a certain threshold. Two types of threshold could have been set – relative and absolute. An absolute one would be defined as an exact amount of income for all states whereas a relative poverty line would have been defined as percentage of a selected economic indicator. Considering the composition of the sample of countries the latter was adopted. The OECD poverty threshold of 50% of median income was used. Application of relative poverty line is considered as advantageous to an absolute threshold as it better reflects the differences in between the considered states. Additional thresholds of 40% and 60% of median income proposed by the LIS were calculated too. These provide the possibility to evaluate the influence of regional price levels for different poverty lines and inquire whether its magnitude changes or not. Considering these poverty lines selection of an appropriate poverty indicator was required.

There exists a variety of indicators useful for measurement of different aspects of poverty. These vary from simple headcounts to poverty gap and income gap ratios. For the purpose of this thesis the headcount ratio calculated by the following identity was adopted. Possibility to simply assess the impacts of regional price levels on incidence of poverty was the main reason for its selection.

$$HCR = \frac{number\ of\ people\ living\ under\ the\ poverty\ threshold}{total\ population}$$

For the purpose of evaluation of impact of spatial price differences both regional price levels adjusted and unadjusted headcount ratios were calculated. The results are provided in Table 9. Figure 11 represents the differences in between the adjusted and unadjusted values graphically. From the graph it is obvious that the adjustment for regional price levels has a significant impact on the measurement of incidence of poverty. However the impact is not the same for all the levels of poverty lines. The absolute effect of adjustments caused by application of regional price levels will be discussed first whereas the dynamics of the changes over the different poverty lines will be discussed latter.

## *6.2 Lowest poverty line - 40% of median income*

The discussion of the impact of regional price levels will be done successively from the lowest to the largest poverty line. Therefore the poverty line of 40% of median income is discussed first. The adjustment for regional price levels led to downward corrections in the majority of cases. Overall the headcount ratios of 52,4% of the states in the sample were adjusted downwards, 42,9% upwards and 4,7% remained unchanged. No effect of the adjustment was encountered only in the case of Germany. Average adjustment was -0,01619. The largest positive adjustments were done to France and Georgia. In the case of France application of regional price levels lead to 0,43% increase in the headcount ratio whereas for Georgia the increase was 0,17%. Even though this result may seem surprising in the case of France it should be considered that France was the state with second largest differences in regional price levels after India. Nonetheless the large adjustment for Georgia is not as surprising as it is a middle income country with significant inter-regional disparities. However the impact of regional price levels is not restricted only to middle income countries as the largest downward adjustments were encountered in the cases of high income countries. The headcount ratio of Hungary underwent the largest downward adjustment of -0,29% amongst the states in the sample. The second largest downward adjustment of -0,28% was done to Italy. At least on this level of the poverty line it seems that the impact of regional price level is the largest for big states with significant inter-regional differences.

### 6.3 Middle poverty line - 50% of median income

Subsequently the impacts of regional price levels on the headcount ratio based on the official OECD poverty line will be assessed. At this level of the poverty threshold the adjustment for regional price levels led to an adjustment for all the considered states. In 57,1% of the cases this was a negative adjustment whereas in 42,9% of the cases it was a positive one. The largest upward adjustment of 0,87% was encountered in the case of Ireland. This is to some extent a surprising result as in the case of the previous poverty line the largest adjustments were done to larger states with significant inter-regional disparities. Considering the size of Ireland and the low differences in between its regional price levels this may indicate a high concentration of individuals close around the poverty line of 50% of median income. Large positive adjustments of 0,81% and 0,21% were done in the case of France and Austria respectively. In the case of Austria this may be explained by the same concentration of individuals along the poverty lines as in the case of Ireland. Significant differences in between the regional price levels are probably the reason for the scale of the adjustment done to France. The largest negative adjustments were encountered in the cases of Italy, Columbia and Mexico of -0,81%, -0,65% and -0,58% respectively.

### 6.4 Highest poverty line - 60% of median income

Finally the impacts of regional price levels on the headcount ratio based on the highest of the poverty lines will be discussed. In the case of the 60% of median poverty threshold no impact of adjustment for regional price levels was found in the case of Austria. For 61,9% of the sample the application of regional price levels led to downward adjustment whereas for 33,3% the adjustment had the opposite direction. The largest negative correction of -1,57% was encountered in the case of Italy. Other significant downward adjustments of -1,53% and -1,15% were made in the case of Columbia and Georgia respectively. Upward corrections of the highest magnitude were encountered in the cases of Ireland and France. Adjustment for regional price levels led to an increase of 1,73% of the headcount ratio in the case of Ireland and 1,34% in the case of France.

Signs of a significant influence of regional price levels on the poverty headcount ratios were found for a variety of states in the dataset. For example in the case of Mexico the adjustment for them led to lower estimates of poverty incidence for all considered levels

of poverty lines. Furthermore in the case of France failure to adjust for regional price levels led to persistent underestimation of the incidence of poverty. However more robust evidence for regional price levels induced overestimation of poverty was found as adjustment for regional price levels led to lower estimates of headcount ratios in the majority of cases for all considered poverty lines. Possible overestimation of poverty incidence was found in the cases of Columbia, Hungary, India, Italy, Mexico, Spain and USA. Two of the four middle income countries included in the dataset were among the states enduring the adjustments of the largest magnitude. The lack of significant adjustments to the state of India might seem surprising considering the differences amongst its regions and the fact that for example Deaton (2003) proved a significant impact of regional price levels on the headcount ratios. Possible explanation for the lack of impact of the adjustment for spatial price differences may be lower concentration of citizens around the considered poverty thresholds. Also there is a possibility that the wrong type of regions was considered as in the previous studies the calculations were done separately for rural and urban parts of the regions. Therefore the rural urban differential has probably higher influence than the differences in between regions. Hence an examination of the impact of spatial price differences in between rural and urban parts of the regions should be done. However such an examination is beyond the scope of this thesis and should be conducted in future research.

# Conclusion

The impact of regional price levels on measures of income inequality and poverty was tested in this thesis. However unavailability of data made this objective harder as only two statistical offices in the world provide regularly regional price levels. Therefore the only possible solution was to estimate them. For that purpose an econometric model was constructed based on already existing estimates of regional price levels for 12 states. After its out-of-sample predictive powers were tested it was used for estimation of regional price levels of 21 states. These regional price levels were then combined with the datasets provided by the LIS database in order to produce both regional price level adjusted and unadjusted Gini coefficients and Headcount ratios. Evidence of persistent bias induced by failure to adjust for differences in regional price levels for both Gini coefficients and headcount ratios was found. This bias was highest for countries outside the WB high income group. The resulting adjustments were also significant for countries with large inter-regional disparities.

The estimation of regional price levels within a sample of 21 countries ranks this thesis among the largest estimations of spatial price differences along with Aten and Heston (2003) and Janský and Kolcunová (2017). Furthermore estimation of regional price levels for a heterogeneous sample of states consisting of both high and middle income countries was conducted only in Aten and Heston (2003). To the best of my knowledge this thesis is also the first work to evaluate the impact of regional price levels on Gini coefficients and headcount ratios for a group of states. It was found that regional price levels can cause distortions in both Gini coefficients and headcount ratios. Considering that these have the biggest impact in middle income countries, organizations responsible for production of comparable price levels such as the ICP should produce regional price levels at least for this group of states as these could have important consequences for the research into distribution of poverty and income inequality. The existence of spatial price differences also ought to be reflected by the respective statistical offices by construction of region specific poverty lines and economic indicators rather than nationwide ones. Even though research into the impacts of regional price levels on the indicators of income inequality and poverty was conducted in this thesis, evaluation of the changes in regional poverty headcounts and distribution of poverty remains for future research. Furthermore the spatial price differences were considered on the level

of defined regions and as was shown for example in the case of India their influence was not as significant. Therefore the impact of rural urban differential in price levels should be addressed in future research too.

# Bibliography

AHEC-ŠONJE, Amina; NESTIĆ, Danijel. International Differences in Price Levels: An Empirical Analysis. *Croatian Economic Survey*, 2002, 4: 85-120

ATEN, Bettina H. Regional Price Parities and Real Regional Income for the United States. *Social Indicators Research*, 2016, 1-21.

ATEN, Bettina H.; FIGUEROA, Eric B.; VENGELEN, B. M. Real personal income and regional price parities for states and metropolitan areas, 2008–2012. change, 2011, 2012: 2012.

ATEN, Betina; MENEZES, Tatiane. Poverty price levels: an application to Brazilian metropolitan areas. In: *World Bank ICP Conference, Washington, DC*. 2002. p. 11-15.

ATEN, Bettina; HESTON, Alan. Regional output differences in international perspective. *Spatial Inequality and Development, UNU-Wider Studies in Development Economics. Oxford University Press, New York*, 2005, 15-36.

BAJGAR, Matěj; JANSKÝ Petr. "Regionální Rozdíly v Kupní Síle: Ceny, Platy, Mzdy a Důchody." Praha: Národohospodářský ústav AV ČR, v. v. i., 2014.

BALASSA, Bela. The purchasing-power parity doctrine: a reappraisal. *Journal of political Economy*, 1964, 72.6: 584-596.

BLIEN, Uwe, et al. Regional price levels and the agglomeration wage differential in western Germany. *The Annals of Regional Science*, 2009, 43.1: 71-88.

BRANDT, Loren; HOLZ, Carsten A. Spatial price differences in China: Estimates and implications. *Economic development and cultural change*, 2006, 55.1: 43-86.

VAN BUUREN, Stef. *Flexible imputation of missing data*. CRC press, 2012.

CADIL, J., MAZOUCH, P., MUSIL, P., & KRAMULOVA, J. (2014). True regional purchasing power: evidence from the Czech Republic. *Post-Communist Economies*, *26*(2), 241-256.

CAI, Fang; DU, Yang; WANG, Meiyan. Demystify the labor statistics in China. *China Economic Journal*, 2013, 6.2-3: 123-133.

COLLIER, Paul. *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press, USA, 2008.

COONDOO, Dipankor; MAJUMDER, Amita; CHATTOPADHYAY, Somnath. Estimating spatial consumer price indices through Engel curve analysis. *Review of Income and Wealth*, 2011, 57.1: 138-155.

DEATON, Angus. Prices and poverty in India, 1987-2000. *Economic and political Weekly*, 2003, 362-368.

DEATON, Angus; HESTON, Alan. Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics*, 2010, 2.4: 1-35.

DECKERS, Thomas; FALK, Armin; SCHILDBERG-HÖRISCH, Hannah. Nominal or real? The impact of regional price levels on satisfaction with life. *The BE Journal of Economic Analysis & Policy*, 2016, 16.3: 1337-1358.

DIKHANOV, Yuri; PALANYANDY, Chellam; CAPILIT, Eileen. Subnational Purchasing Power Parities toward Integration of International Comparison Program and Consumer Price Index: The Case of the Philippines. 2011.

FENG, Shuaizhang; HU, Yingyao; MOFFITT, Robert. *Long run trends in unemployment and labor force participation in China*. National Bureau of Economic Research, 2015.

GADDIS, Isis. Prices for poverty analysis in Africa. *Browser Download This Paper*, 2016.

GELB, Alan; DIOFASI, Anna. What Determines Purchasing Power Parity Exchange Rates?. 2015.

GILES, John; ALBERT, Park; ZHANG, Juwei. What is China's true unemployment rate?. *China Economic Review*, 2005, 16.2: 149-170.

HESTON, Alan; NUXOLL, Daniel A.; SUMMERS, Robert. The Differential-Productivity Hypothesis and Purchasing-Power Parties: Some New Evidence. *Review of International Economics*, 1994, 2.3: 227-243.

JANSKÝ, Petr; KOLCUNOVÁ, Dominika. Regional differences in price levels across the European Union and their implications for its regional policy. *The Annals of Regional Science*, 2017, 58.3: 641-660.

KRAMULOVÁ, Jana; MUSIL, Petr; ZEMAN, Jan; MICHLOVÁ, Radka. 2016. "Regional Price Levels in the Czech Republic - Past and Current Perspectives."

KRAVIS, Irving; LIPSEY, Robert E. National price levels and the prices of tradables and nontradables. 1988.

KRUGMAN, Paul. Increasing returns and economic geography. *Journal of political economy*, 1991, 99.3: 483-499.

LI, Xiumin; ZHANG, Lili; DU, Yashu. Study on the method of regional purchasing power parity in China. *China-USA Business Review*, 2005, 4.7: 41-48.

LIU, Qian. Unemployment and labor force participation in urban China. *China Economic Review*, 2012, 23.1: 18-33.

CHAKRABARTY, Manisha; MAJUMDER, Amita; RAY, Ranjan. Preferences, spatial prices and inequality. *The Journal of Development Studies*, 2015, 51.11: 1488-1501.

MAJUMDER, Amita, et al. *Estimates of Spatial Prices in India and their Sensitivity to Alternative Estimation Methods and Choice of Items*. Monash University, Department of Economics, 2015.

MAJUMDER, Amita, et al. Estimating intra country and cross country purchasing power parities from household expenditure data using single equation and complete demand systems approach: India and Vietnam. *Economics discussion paper*, 2011, 34.11.

MISHRA, Ankita; RAY, Ranjan. Spatial variation in prices and expenditure inequalities in Australia. *Economic Record*, 2014, 90.289: 137-159.

PITTAU, Maria Grazia; ZELLI, Roberto; MASSARI, Riccardo. Do spatial price indices reshuffle the Italian income distribution?. *Modern Economy*, 2011, 2.03: 259.

RADVANSKY, Marek, et al. *Computing real income at NUTS 3 regions*. EcoMod, 2012.

"Reale Kaukraft 2008: Einkommen Unter Berücksichtigung Des Regionalen Preisniveaus." 2009. Österrechische Gesellshaft für Marketing.

ROKICKI, Bartlomiej; HEWINGS, Geoffrey JD. Regional convergence within particular country—An approach based on the regional price deflators. Economic Modelling, 2016, 57: 171-179.

ROOS, Michael WM. Regional price levels in Germany. *Applied Economics*, 2006, 38.13: 1553-1566.

SAMUELSON, Paul A. Theoretical notes on trade problems. *The Review of Economics and Statistics*, 1964, 145-154.

STRÖHL, Gerd. Zwischenörtlicher Vergleich des Verbraucherpreisniveaus in 50 Städten. *Wirtschaft und Statistik*, 1994, 6.1994: 415-434.

SUEDEKUM, Jens. Agglomeration and regional costs of living. *Journal of Regional Science*, 2006, 46.3: 529-543.

SUMMERS, Robert. International price comparisons based upon incomplete data. *Review of Income and Wealth*, 1973, 19.1: 1-16.

SUMNER, Andy. Global Poverty and the New Bottom Billion: What if Three-Quarters of the World's Poor Live in Middle-income Countries?. *IDS Working Papers*, 2010, 2010.349: 01-43.

"UK Relative Regional Consumer Price Levels for Goods and Services for 2010." 2011. Office for National Statistics. https://data.gov.uk/dataset/regional_consumer_price_levels.

# List of Abbreviations

| | |
|---|---|
| ADB: | Asian Development Bank |
| BEA: | Bureau of Economic Analysis |
| BLS: | Bureau of Labor Statistic |
| CPD method: | Country Product Dummy method |
| CPI: | Consumer Price Index |
| EKS method: | Èltetö-Köves-Szulc method |
| EU: | European Union |
| GDP: | Gross Domestic Product |
| ICP: | International Comparison Program |
| LIS: | Luxembourg Income Study |
| NBS: | National Bureau of Statistics (China) |
| NSS: | National Sample Survey |
| OECD: | Organization for Economic Cooperation and Development |
| ONS: | Office for National Statistics |
| PSA: | Philippine Statistical Authority |
| PSY: | Philippine Statistics Yearbook |
| PPP: | Purchasing Power Parity |
| PWT: | Penn World Tables |
| RCPL: | Regional Consumer Price Level |
| Turkstat: | Turkish Statistical Institute |
| UK: | United Kingdom |
| UNESCO: | United Nations Educational, Scientific and Cultural Organization |
| USA: | United States of America |
| WB: | World Bank |

# List of Appendices

# Appendices

### Appendix 1: Table 1 - Sources of Regional Price Levels

| State | Year | Source |
|---|---|---|
| **Czech Republic** | *2012* | Estimation based on adjusted Eurostat methodology and consumption data.<br><br>*Source: KRAMULOVÁ, Jana; MUSIL, Petr; ZEMAN, Jan; MICHLOVÁ, Radka. 2016. "Regional Price Levels in the Czech Republic - Past and Current Perspectives."* |
| **Poland** | *2011* | Data set constructed for the calculation of real regional income.<br><br>*Source: ROKICKI, Bartlomiej; HEWINGS, Geoffrey JD. Regional convergence within particular country—An approach based on the regional price deflators. Economic Modelling, 2016, 57: 171-179.* |
| **Austria** | *2008* | Joint paper of the Austrian Statistical Office and Österreichische Gessellschaft für Marketing.<br><br>*Source: "Reale Kaukraft 2008: Einkommen Unter Berücksichtigung Des Regionalen Preisniveaus" (2009)* |
| **Slovakia** | *2009* | Estimation constructed on family budget survey data.<br><br>*Source: RADVANSKY, Marek, et al. Computing real income at NUTS 3 regions. EcoMod, 2012.* |
| **Germany** | *2010* | Regional deflators based on a prediction model constructed on data for 50 German cities. 1993 indices shifted to 2010 by regional CPI.<br><br>*Source: ROOS, Michael WM. Regional price levels in Germany. Applied Economics, 2006, 38.13: 1553-1566.* |
| **United Kingdom** | *2010* | Data produced by the Office for National Statistics during the construction of the UK Spatial Adjustment Factors for Eurostat.<br><br>*Source: "UK Relative Regional Consumer Price Levels for Goods and Services for 2010." 2011. Office for National Statistics.* https://data.gov.uk/dataset/regional_consumer_price_levels . |
| **Italy** | *2006* | Estimates produced by the National Bank of Italy in a joint project with Italian Office of Statistics.<br><br>*Source: PITTAU, Maria Grazia; ZELLI, Roberto; MASSARI, Riccardo. Do spatial price indices reshuffle the Italian income distribution?. Modern Economy, 2011, 2.03: 259.* |
| **USA** | *2014* | Price levels constructed by the BEA and BLS, based on consumption data, following the methodology outlined in Aten (2015).<br><br>*Source:* "Regional Data." 2017. *Bureau of Economic Analysis*. February 20. https://www.bea.gov/iTable/iTable.cfm?reqid=70&step=1&isuri=1&acrdn=8#reqid=70&step=30&isuri=1&7022=101&7023=8&7024=non- |

| | | |
|---|---|---|
| | | industry&7033=-1&7025=0&7026=xx&7027=2014&7001=8101&7028=-1&7031=0&7040=-1&7083=levels&7029=101&7090=70. |
| **Australia** | *2009* | Estimation done by the the Exact Affine Stone Index demand system, based on the data from the Household Expenditure Survey.<br><br>*Source: MISHRA, Ankita; RAY, Ranjan. Spatial variation in prices and expenditure inequalities in Australia. Economic Record, 2014, 90.289: 137-159.* |
| **Turkey** | *2014* | Data constructed by Turkstat as a by-product of the construction of the Spatial Adjustment Factor, based on a 2012 survey and shifted to 2014 by regional CPI.<br><br>*Source: "Purchasing Power Parity (PPP)." 2017. Turkish Statistical Institute. http://www.turkstat.gov.tr/PreTablo.do?alt_id=1065.* |
| **Philippines** | *2010* | Data set constructed by a project of the ADB, based on data collected for the construction of CPI<br><br>*Source: DIKHANOV, Yuri; PALANYANDY, Chellam; CAPILIT, Eileen. Subnational Purchasing Power Parities toward Integration of International Comparison Program and Consumer Price Index: The Case of the Philippines. 2011.* |
| **China** | *2014* | Regional deflators based on a 1990 basket published by the NBS. Shifted by CPI.<br><br>*Source: BRANDT, Loren; HOLZ, Carsten A. Spatial price differences in China: Estimates and implications. Economic development and cultural change, 2006, 55.1: 43-86.* |
| | *2002* | Regional indices based on 127 commodities. (Shanghai as numeraire)<br><br>*Source: LI, Xiumin; ZHANG, Lili; DU, Yashu. Study on the method of regional purchasing power parity in China. China-USA Business Review, 2005, 4.7: 41-48.* |

**Appendix 2: Table 2 - Explanatory variables**

| Variable | Description | Source |
|---|---|---|
| **Increl** | Disposable income per head.<br>(relative value) | 1, 2, 3, 6, 7 |
| **GDPrel** | GDP per capita.<br>(relative value) | 1, 2, 6, 7 |
| **Emprel** | Employment rate<br>(relative value) | 1, 2, 6, 7 |
| **Unemprel** | Regional rate of unemployment of people aged 15 and over.<br>(relative value) | 1, 2, 3, 4, 5, 6, 7 |
| **Partrate** | Participation rate<br>(relative value) | 1,2, 6, 7 |
| **Popdensrel** | Regional population density per kilometre squared.<br>(relative value) | 1, 2, 6, 7 |
| **Areaperc** | Percentage of total country area that the region covers. | 1, 2, 6, 7 |
| **Popperc** | Percentage of the population living within the given region. | 1, 2, 6, 7 |
| **Capital** | Indicates the presence of capital in the region.<br>(Dummy) | |
| **UNESCO** | Indicates the presence of a monument from the UNESCO world heritage list.<br>(Dummy) | |
| **City1** | Indicates the presence of a city with more than 1% of total country population.<br>(Dummy) | |
| **City2** | Indicates the presence of a city with more than 2% of total country population.<br>(Dummy) | |
| **Sea** | Indicates that the region has access to sea.<br>(Dummy) | |
| **Airp** | Indicates presence of an international airport in the region.<br>(Dummy) | |

1: OECD 2: Eurostat 3: Turkstat 4: Czech statistical office 5: Slovak statistical office 6: Philippine Statistical Yearbook 7: National Bureau of Statistics (China)

**Appendix 3: Table 5 - Description of the LIS dataset**

| State | Regions | Number of Observations | Regional Price Level |
|---|---|---|---|
| Australia | AU1: New South Wales | 3 314 | 97,68452 |
| Year: 2010 | AU2: Victoria | 3 106 | 95,48913 |
| | AU3: Queensland | 2 703 | 99,89687 |
| | AU4: South Australia | 2 800 | 97,03205 |
| | AU5: Western Australia | 2 744 | 104,37522 |
| | AU6: Tasmania | 1 864 | 93,18972 |
| | AU8: Australian Capital Territory | 1 540 | 118,92521 |
| | Missing | 0 | |
| Austria | AT11: Burgenland (AT) | 207 | 97,78879 |
| Year: 2010 | AT12: Lower Austria | 1 149 | 102,96596 |
| | AT13: Vienna | 1 315 | 103,42648 |
| | AT21: Carinthia | 417 | 98,74733 |
| | AT22: Styria | 828 | 100,79831 |
| | AT31: Upper Austria | 1 078 | 100,37497 |
| | AT32: Salzburg | 425 | 99,7607 |
| | AT33: Tyrol | 517 | 100,48559 |
| | AT34: Vorarlberg | 251 | 97,69743 |
| | Missing | 0 | |
| Canada | CA10: Newfoundland and Labrador | 1 093 | 94,06136 |
| Year: 2010 | CA11: Prince Edward Island | 685 | 93,37394 |
| | CA12: Nova Scotia | 1 516 | 94,4721 |
| | CA13: New Brunswick | 1 475 | 94,63277 |
| | CA24: Quebec | 4 888 | 98,28488 |
| | CA35: Ontario | 6 764 | 99,11353 |
| | CA46: Manitoba | 1 705 | 97,24282 |
| | CA47: Saskatchewan | 1 804 | 98,98334 |
| | CA48: Alberta | 2 474 | 102,42638 |
| | CA59: British Columbia | 2 422 | 99,25471 |
| | Missing | 193 | |
| Czech Republic | CZ010: Prague | 871 | 115,82433 |
| Year: 2010 | CZ020: Central Bohemia | 1 003 | 103,21719 |
| | CZ031: South Bohemia | 630 | 100,10262 |
| | CZ032: Plzen | 476 | 99,95109 |
| | CZ041: Karlovy Vary | 214 | 95,20527 |
| | CZ042: Ústí nad Labem | 720 | 95,9113 |
| | CZ051: Liberec | 348 | 97,71293 |
| | CZ052: Hradec Králové | 460 | 98,43889 |
| | CZ053: Pardubice | 419 | 97,5612 |
| | CZ063: Vysocina | 486 | 98,55572 |
| | CZ064: South Moravia | 955 | 99,17213 |

| | | | |
|---|---|---|---|
| | CZ071: Olomouc | 536 | 96,64227 |
| | CZ072: Zlín | 536 | 96,83851 |
| | CZ080: Moravia-Silesia | 1 212 | 96,8367 |
| | Missing | 0 | |
| Germany | DE1: Baden-Württemberg | 1 413 | 102,64479 |
| | DE2: Bavaria | 1 846 | 105,19746 |
| | DE3: Berlin | 513 | 100,64233 |
| | DE4: Brandenburg | 524 | 96,49666 |
| | DE5: Bremen | 81 | 99,60815 |
| | DE6: Hamburg | 198 | 103,8689 |
| | DE7: Hesse | 833 | 100,49072 |
| | DE8: Mecklenburg-Vorpommern | 318 | 94,42141 |
| | DE9: Lower Saxony | 1 093 | 100,30331 |
| Year: 2010 | DEA: North Rhine-Westphalia | 2 367 | 100,48073 |
| | DEB: Rhineland-Palatinate | 573 | 100,15773 |
| | DEC: Saarland | 138 | 97,46461 |
| | DED: Saxony | 809 | 95,42395 |
| | DEE: Saxony-Anhalt | 509 | 94,7704 |
| | DEF: Schleswig-Holstein | 407 | 99,07944 |
| | DEG: Thuringia | 524 | 95,75457 |
| | Missing | 0 | |
| Greece | EL30: Attica | 1 442 | 100,75007 |
| | EL41: North Aegean | 170 | 98,987 |
| | EL42: South Aegean | 158 | 98,65114 |
| | EL43: Crete | 378 | 96,51071 |
| | EL51: Eastern Macedonia, Thrace | 459 | 97,10792 |
| | EL52: Central Macedonia | 1 109 | 99,38731 |
| | EL53: Western Macedonia | 217 | 97,95745 |
| Year: 2010 | EL54: Epirus | 219 | 98,30731 |
| | EL61: Thessaly | 503 | 98,97689 |
| | EL62: Ionian Islands | 92 | 96,77962 |
| | EL63: Western Greece | 536 | 97,67833 |
| | EL64: Central Greece | 170 | 98,12658 |
| | EL65: Peloponnese | 401 | 99,72298 |
| | Missing | 0 | |
| Italy | ITC1: Piedmont | 711 | 102,29369 |
| | ITC2: Aosta Valley | 46 | 100,78736 |
| | ITC3: Liguria | 311 | 101,79711 |
| | ITC4: Lombardy | 803 | 106,29012 |
| | ITF1: Abruzzo | 202 | 96,56418 |
| Year: 2010 | ITF2: Molise | 116 | 94,74972 |
| | ITF3: Campania | 752 | 94,57011 |
| | ITF4: Apulia | 454 | 93,88829 |
| | ITF5: Basilicata | 126 | 92,29205 |
| | ITF6: Calabria | 196 | 93,28838 |

| | | | |
|---|---|---|---|
| | ITG1: Sicily | 587 | 93,83421 |
| | ITG2: Sardinia | 342 | 93,98694 |
| | ITH2: Province of Trento | 172 | 102,19652 |
| | ITH3: Veneto | 512 | 101,99643 |
| | ITH4: Friuli-Venezia Giulia | 214 | 101,14435 |
| | ITH5: Emilia-Romagna | 708 | 104,06963 |
| | ITI1: Tuscany | 615 | 101,61602 |
| | ITI2: Umbria | 277 | 99,05767 |
| | ITI3: Marche | 355 | 99,55778 |
| | ITI4: Lazio | 452 | 101,37973 |
| | Missing | 0 | |
| Mexico | ME01: Aguascalientes | 329 | 98,42106 |
| Year: 2010 | ME02: Baja California Norte | 504 | 105,45766 |
| | ME03: Baja California Sur | 299 | 104,64658 |
| | ME04: Campeche | 427 | 98,1851 |
| | ME05: Coahuila | 564 | 99,82382 |
| | ME06: Colima | 361 | 99,802 |
| | ME07: Chiapas | 2 805 | 92,96052 |
| | ME08: Chihuahua | 836 | 98,88823 |
| | ME09: Federal District (MX) | 2 799 | 108,86124 |
| | ME10: Durango | 497 | 95,31053 |
| | ME11: Guanajuato | 1 901 | 94,00942 |
| | ME12: Guerrero | 868 | 94,75465 |
| | ME13: Hidalgo | 513 | 93,66464 |
| | ME14: Jalisco | 595 | 99,50776 |
| | ME15: Mexico | 2 748 | 96,56129 |
| | ME16: Michoacan | 700 | 94,88315 |
| | ME17: Morelos | 462 | 96,32873 |
| | ME18: Nayarit | 408 | 98,33079 |
| | ME19: Nuevo Leon | 411 | 109,40972 |
| | ME20: Oaxaca | 1 051 | 94,72115 |
| | ME21: Puebla | 662 | 94,22843 |
| | ME22: Queretaro | 452 | 96,89616 |
| | ME23: Quintana Roo | 350 | 102,93532 |
| | ME24: San Luis Potosi | 539 | 95,17982 |
| | ME25: Sinaloa | 458 | 98,48497 |
| | ME26: Sonora | 649 | 100,33726 |
| | ME27: Tabasco | 487 | 92,58976 |
| | ME28: Tamaulipas | 558 | 97,29986 |
| | ME29: Tlaxcala | 380 | 91,16649 |
| | ME30: Veracruz | 898 | 96,10716 |
| | ME31: Yucatan | 2 719 | 97,01691 |
| | ME32: Zacatecas | 425 | 93,86371 |
| | Missing | 0 | |
| Poland | PL11: Lodzkie | 2 680 | 100,38766 |

| | | | |
|---|---|---|---|
| Year: 2010 | PL12: Mazovia | 5 388 | 104,72956 |
| | PL21: Lesser Poland | 3 181 | 99,52277 |
| | PL22: Silesia | 4 504 | 103,93282 |
| | PL31: Lublin Province | 2 186 | 97,71716 |
| | PL32: Podkarpacia | 1 956 | 95,84125 |
| | PL33: Swietokrzyskie | 1 285 | 96,56056 |
| | PL34: Podlasie | 1 204 | 96,91237 |
| | PL41: Greater Poland | 3 143 | 101,39063 |
| | PL42: West Pomerania | 1 617 | 98,79538 |
| | PL43: Lubusz | 963 | 97,72706 |
| | PL51: Lower Silesia | 2 950 | 99,97349 |
| | PL52: Opole region | 966 | 97,84054 |
| | PL61: Kuyavian-Pomerania | 1 976 | 98,02952 |
| | PL62: Warmian-Masuria | 1 381 | 97,73204 |
| | PL63: Pomerania | 2 032 | 99,7121 |
| | Missing | 0 | |
| Slovakia | SK010: Bratislava Region | 464 | 111,71823 |
| Year: 2010 | SK021: Trnava Region | 556 | 100,4198 |
| | SK022: Trencín Region | 655 | 100,37149 |
| | SK023: Nitra Region | 679 | 100,4306 |
| | SK031: Žilina Region | 677 | 99,978 |
| | SK032: Banská Bystrica Region | 691 | 99,05364 |
| | SK041: Prešov Region | 733 | 97,85512 |
| | SK042: Košice Region | 745 | 97,06646 |
| | Missing | 0 | |
| Spain | ES11: Galicia | 879 | 97,38213 |
| Year: 2010 | ES12: Asturias | 566 | 98,08939 |
| | ES13: Cantabria | 416 | 97,51112 |
| | ES21: Basque Country | 747 | 104,61034 |
| | ES22: Navarra | 434 | 103,11496 |
| | ES23: La Rioja | 447 | 98,16676 |
| | ES24: Aragon | 612 | 100,95003 |
| | ES30: Madrid | 1 238 | 102,1206 |
| | ES41: Castile and León | 881 | 101,4724 |
| | ES42: Castile-La Mancha | 708 | 97,72097 |
| | ES43: Extremadura | 496 | 94,58837 |
| | ES51: Catalonia | 1 465 | 101,16768 |
| | ES52: Valencia | 1 025 | 95,5291 |
| | ES53: Balearic Islands | 387 | 96,73 |
| | ES61: Andalusia | 1 471 | 96,5266 |
| | ES62: Murcia | 488 | 94,00797 |
| | ES63: Ceuta | 113 | 94,47599 |
| | ES64: Melilla | 112 | 93,62811 |
| | ES70: Canary Islands | 624 | 93,26133 |
| | Missing | 0 | |

| | | | |
|---|---|---|---|
| Switzerland | CH01: Lake Geneva Region | 1 266 | 99,535 |
| Year: 2010 | CH02: Espace Mittelland | 1 856 | 101,3736 |
| | CH03: Northwestern Switzerland | 1 065 | 98,51578 |
| | CH04: Zürich | 1 374 | 101,24393 |
| | CH05: Eastern Switzerland | 987 | 103,073 |
| | CH06: Central Switzerland | 688 | 100,64578 |
| | CH07: Ticino | 266 | 95,62398 |
| | Missing | 0 | |
| UK | UKC: North East England | 1 025 | 95,44926 |
| Year: 2010 | UKD: North West England | 2 627 | 97,48468 |
| | UKE: Yorkshire and The Humber | 1 920 | 96,63182 |
| | UKF: East Midlands | 1 637 | 97,61285 |
| | UKG: West Midlands | 1 947 | 96,7505 |
| | UKH: East of England | 2 117 | 100,85186 |
| | UKI: Greater London | 2 205 | 111,18624 |
| | UKJ: South East England | 2 904 | 102,98701 |
| | UKK: South West England | 1 773 | 100,76429 |
| | UKL: Wales | 1 174 | 96,82547 |
| | UKM: Scotland | 4 126 | 103,83209 |
| | UKN: Northern Ireland | 1 895 | 96,42835 |
| | Missing | 0 | |
| USA | US01: Alabama | 844 | 94,20918 |
| Year: 2010 | US02: Alaska | 982 | 102,09312 |
| | US04: Arizona | 977 | 94,20788 |
| | US05: Arkansas | 754 | 94,09685 |
| | US06: California | 6 555 | 97,38018 |
| | US08: Colorado | 1 643 | 96,62006 |
| | US09: Connecticut | 1 630 | 108,34795 |
| | US10: Delaware | 1 184 | 98,32489 |
| | US11: District of Columbia | 1 333 | 136,64132 |
| | US12: Florida | 3 161 | 96,59215 |
| | US13: Georgia | 1 674 | 94,56267 |
| | US15: Hawaii | 1 180 | 98,70772 |
| | US16: Idaho | 766 | 93,94803 |
| | US17: Illinois | 2 295 | 97,35515 |
| | US18: Indiana | 1 120 | 94,76378 |
| | US19: Iowa | 1 338 | 97,03447 |
| | US20: Kansas | 1 105 | 96,9609 |
| | US21: Kentucky | 1 058 | 93,96895 |
| | US22: Louisiana | 743 | 96,43477 |
| | US23: Maine | 1 269 | 96,0501 |
| | US24: Maryland | 1 799 | 102,50412 |
| | US25: Massachusetts | 1 108 | 104,22354 |
| | US26: Michigan | 1 697 | 94,25994 |
| | US27: Minnesota | 1 728 | 97,92741 |

| | | | |
|---|---|---|---|
| | US28: Mississippi | 765 | 93,01775 |
| | US29: Missouri | 1 243 | 95,53213 |
| | US30: Montana (US) | 684 | 95,63269 |
| | US31: Nebraska | 1 116 | 98,43274 |
| | US32: Nevada | 1 172 | 93,77209 |
| | US33: New Hampshire | 1 389 | 101,31379 |
| | US34: New Jersey | 1 511 | 104,4675 |
| | US35: New Mexico | 738 | 94,65341 |
| | US36: New York | 3 360 | 100,35205 |
| | US37: North Carolina | 1 499 | 94,94507 |
| | US38: North Dakota | 931 | 100,06222 |
| | US39: Ohio | 2 019 | 95,61473 |
| | US40: Oklahoma | 981 | 96,04995 |
| | US41: Oregon | 1 026 | 94,31855 |
| | US42: Pennsylvania | 2 231 | 98,23191 |
| | US44: Rhode Island | 1 221 | 99,84308 |
| | US45: South Carolina | 991 | 93,65945 |
| | US46: South Dakota | 1 171 | 98,86895 |
| | US47: Tennessee | 992 | 95,64637 |
| | US48: Texas | 4 238 | 97,02147 |
| | US49: Utah | 798 | 94,14224 |
| | US50: Vermont | 1 016 | 98,20589 |
| | US51: Virginia | 1 614 | 99,56412 |
| | US53: Washington | 1 285 | 97,56713 |
| | US54: West Virginia | 787 | 94,06694 |
| | US55: Wisconsin | 1 436 | 96,378 |
| | US56: Wyoming | 1 031 | 99,76968 |
| | Missing | 0 | |
| India | IN01: Jammu and Kashmir | 720 | 99,22028 |
| Year: 2011 | IN02: Himachal Pradesh | 1 476 | 102,88209 |
| | IN03: National Capital Territory of Delhi | 1 702 | 115,97074 |
| | IN04: Rajasthan | 85 | 97,53693 |
| | IN05: Uttar Pradesh | 468 | 94,66474 |
| | IN06: Sikkim | 1 806 | 105,12953 |
| | IN07: Arunachal Pradesh | 899 | 123,72837 |
| | IN08: Nagaland | 2 707 | 89,98206 |
| | IN09: Meghalaya | 3 824 | 99,16251 |
| | IN10: Assam | 1 547 | 94,26223 |
| | IN11: West Bengal | 107 | 94,97087 |
| | IN12: Gujarat | 159 | 98,25063 |
| | IN13: Dadra & Nagar Haveli | 110 | 96,04156 |
| | IN14: Maharashtra | 88 | 97,74852 |
| | IN15: Daman & Diu | 78 | 102,43362 |
| | IN17: Kerala | 220 | 96,47694 |
| | IN18: Punjab | 134 | 101,11092 |

| | | | |
|---|---|---|---|
| | IN19: Chandigarh | 991 | 164,86825 |
| | IN20: Haryana | 2 435 | 97,76857 |
| | IN21: Uttaranchal | 853 | 94,5264 |
| | IN22: Bihar | 2 058 | 92,67257 |
| | IN23: Jharkhand | 1 324 | 92,97467 |
| | IN24: Manipur | 3 123 | 97,5048 |
| | IN25: Mizoram | 1 895 | 109,0756 |
| | IN26: Tripura | 59 | 85,03115 |
| | IN27: Orissa | 60 | 93,26485 |
| | IN28: Madhya Pradesh | 3 309 | 95,55097 |
| | IN29: Chhattisgarh | 2 203 | 93,89714 |
| | IN30: Andhra Pradesh | 3 865 | 95,0646 |
| | IN31: Karnataka | 188 | 96,77575 |
| | IN32: Goa | 1 570 | 95,3345 |
| | IN33: Tamil Nadu | 1 982 | 97,32202 |
| | IN34: Puducherry | 107 | 105,42793 |
| | Missing | 0 | |
| Ireland | IE011: Border | 428 | 100,02819 |
| Year: 2010 | IE012: Midlands | 300 | 97,52517 |
| | IE013: West | 447 | 100,93112 |
| | IE021: Dublin | 963 | 106,53841 |
| | IE022: Mid-East | 426 | 99,81379 |
| | IE023: Mid-West | 468 | 99,3714 |
| | IE024: South-East (IE) | 469 | 98,9148 |
| | IE025: South-West (IE) | 832 | 101,40638 |
| | Missing | 0 | |
| Hungary | HU101: Budapest | 410 | 123,476 |
| Year: 2012 | HU102: Pest | 201 | 98,49111 |
| | HU211: Fejér | 57 | 98,93956 |
| | HU212: Komárom-Esztergom | 60 | 101,92149 |
| | HU213: Veszprém | 101 | 97,31632 |
| | HU221: Gyor-Moson-Sopron | 74 | 100,89319 |
| | HU222: Vas | 33 | 97,5122 |
| | HU223: Zala | 90 | 96,89514 |
| | HU231: Baranya | 88 | 94,58879 |
| | HU232: Somogy | 76 | 97,64167 |
| | HU233: Tolna | 32 | 95,94272 |
| | HU311: Borsod-Abaúj-Zemplén | 167 | 98,18964 |
| | HU312: Heves | 20 | 100,66321 |
| | HU313: Nógrád | 66 | 97,68698 |
| | HU321: Hajdú-Bihar | 148 | 98,47484 |
| | HU322: Jász-Nagykun-Szolnok | 40 | 94,30388 |
| | HU323: Szabolcs-Szatmár-Bereg | 118 | 94,82616 |
| | HU331: Bács-Kiskun | 140 | 98,19402 |
| | HU332: Békés | 73 | 96,41258 |

| | | | |
|---|---|---|---|
| | HU333: Csongrád | 67 | 98,55181 |
| | Missing | 0 | |
| Colombia | CO05: Antioquia | 1 139 | 102,34922 |
| Year: 2010 | CO08: Atlántico | 861 | 96,40002 |
| | CO11: Bogotá Capital District | 871 | 125,72122 |
| | CO13: Bolívar | 669 | 94,8765 |
| | CO15: Boyacá | 822 | 99,88452 |
| | CO17: Caldas | 665 | 110,18948 |
| | CO18: Caquetá | 815 | 97,37713 |
| | CO19: Cauca | 850 | 94,6947 |
| | CO20: Cesar | 770 | 95,06856 |
| | CO23: Córdoba (CO) | 778 | 95,94242 |
| | CO25: Cundinamarca | 381 | 96,06268 |
| | CO27: Chocó | 738 | 95,2001 |
| | CO41: Huila | 805 | 100,72779 |
| | CO44: La Guajira | 706 | 99,96207 |
| | CO47: Magdalena | 690 | 95,87624 |
| | CO50: Meta | 634 | 102,67151 |
| | CO52: Nariño | 780 | 91,43467 |
| | CO54: Norte de Santander | 727 | 92,57247 |
| | CO63: Quindio | 869 | 95,62277 |
| | CO66: Risaralda | 813 | 103,5692 |
| | CO68: Santander | 833 | 103,37885 |
| | CO70: Sucre | 777 | 95,02039 |
| | CO73: Tolima | 725 | 99,99614 |
| | CO76: Valle del Cauca | 908 | 96,12572 |
| | Missing | 0 | |
| France | Île de France | 1 625 | 159,17622 |
| Year: 2010 | Bassin Parisien | 1 922 | 129,01237 |
| | Nord - Pas-de-Calais | 761 | 94,90316 |
| | Est (FR) | 936 | 100,16957 |
| | Ouest (FR) | 1 568 | 114,48458 |
| | Sud-Ouest (FR) | 1 199 | 108,10981 |
| | Centre-Est (FR) | 1 138 | 111,25672 |
| | Méditerranée | 1 193 | 110,44199 |
| | Guadeloupe | 1 124 | 88,44145 |
| | Martinique | 1 026 | 88,86185 |
| | Guyane | 956 | 90,34394 |
| | La Réunion | 1 169 | 87,09877 |
| | Mayotte | 1 180 | 90,30203 |
| | Missing | 0 | |
| Denmark | DK011: City of Copenhagen | 12 003 | 103,79266 |
| Year: 2010 | DK012: Copenhagen suburbs | 8 035 | 100,1818 |
| | DK013: North Zealand | 6 363 | 99,79058 |
| | DK014: Bornholm | 654 | 96,15402 |

| | | | |
|---|---|---|---|
| | DK021: East Zealand | 3 378 | 98,19877 |
| | DK022: West and South Zealand | 8 922 | 99,52144 |
| | DK031: Fyn | 7 532 | 97,39823 |
| | DK032: South Jutland | 10 823 | 101,35144 |
| | DK041: West Jutland | 12 673 | 100,50311 |
| | DK042: East Jutland | 6 265 | 100,30124 |
| | DK050: North Jutland | 8 997 | 100,2151 |
| | Missing | 0 | |
| Georgia | Kakheti | 682 | 96,96171 |
| Year: 2010 | Tbilisi | 1 051 | 149,99953 |
| | Shida Kartli | 443 | 92,89632 |
| | Kvemo Kartli | 603 | 95,63793 |
| | Adjara (Automomous Region) | 450 | 92,35578 |
| | Samegrelo-Zemo Svaneti | 509 | 96,88451 |
| | Imereti, Racha-Lechkhumi and Kvemo Svaneti | 936 | 109,31324 |
| | Other regions | 995 | 100,10113 |
| | Missing | 0 | |

**Appendix 4: Table 6 - Sources of data for prediction**

| State | Source |
|---|---|
| **Australia** | 1 |
| **Austria** | 1, 2 |
| **Canada** | 1 |
| **Colombia** | 1, 3 |
| **Czech Republic** | 1, 2 |
| **Denmark** | 1, 5 |
| **France** | 2 |
| **Georgia** | 6 |
| **Germany** | 1, 2 |
| **Greece** | 1 |
| **Hungary** | 1, 3 |
| **India** | 1, 3 |
| **Ireland** | 1, 2, 4 |
| **Italy** | 1 |
| **Mexico** | 1 |
| **Poland** | 1 |
| **Slovakia** | 1 |
| **Spain** | 1 |
| **Switzerland** | 1 |
| **UK** | 1 |
| **USA** | 1 |

1: OECD 2: Eurostat 3: LIS 4: Central Statistics Office (Ireland) 5: Statistics Denmark 6: National Statistics Office of Georgia

**Appendix 5: Table 7 – Gini coefficients**

| | Household income | | Per capita icome | | Equalised income | |
|---|---|---|---|---|---|---|
| **State** | **Adjusted** | **Unadjusted** | **Adjusted** | **Unadjusted** | **Adjusted** | **Unadjusted** |
| **Australia** | 0,389 | 0,39 | 0,347 | 0,348 | 0,332 | 0,334 |
| **Austria** | 0,353 | 0,353 | 0,303 | 0,304 | 0,28 | 0,28 |
| **Canada** | 0,363 | 0,364 | 0,321 | 0,322 | 0,301 | 0,302 |
| **Colombia** | 0,51 | 0,523 | 0,53 | 0,543 | 0,496 | 0,511 |
| **Czech Republic** | 0,319 | 0,323 | 0,259 | 0,266 | 0,251 | 0,256 |
| **Denmark** | 0,347 | 0,347 | 0,262 | 0,263 | 0,255 | 0,254 |
| **France** | 0,342 | 0,342 | 0,322 | 0,323 | 0,292 | 0,292 |
| **Georgia** | 0,466 | 0,479 | 0,436 | 0,457 | 0,427 | 0,446 |
| **Germany** | 0,353 | 0,355 | 0,303 | 0,305 | 0,286 | 0,288 |
| **Greece** | 0,367 | 0,367 | 0,337 | 0,338 | 0,327 | 0,328 |
| **Hungary** | 0,343 | 0,35 | 0,301 | 0,317 | 0,281 | 0,293 |
| **India** | 0,52 | 0,524 | 0,511 | 0,515 | 0,499 | 0,503 |
| **Ireland** | 0,351 | 0,351 | 0,323 | 0,32 | 0,298 | 0,296 |
| **Italy** | 0,357 | 0,362 | 0,348 | 0,355 | 0,323 | 0,331 |
| **Mexico** | 0,474 | 0,479 | 0,493 | 0,499 | 0,459 | 0,465 |
| **Poland** | 0,36 | 0,362 | 0,344 | 0,346 | 0,312 | 0,314 |
| **Slovakia** | 0,343 | 0,344 | 0,27 | 0,274 | 0,262 | 0,264 |
| **Spain** | 0,37 | 0,372 | 0,349 | 0,352 | 0,335 | 0,337 |
| **Switzerland** | 0,348 | 0,348 | 0,33 | 0,329 | 0,299 | 0,299 |
| **UK** | 0,383 | 0,387 | 0,352 | 0,356 | 0,335 | 0,339 |
| *based on ONS RCPL* | 0,384 | | 0,353 | | 0,336 | |
| **USA** | 0,414 | 0,415 | 0,403 | 0,404 | 0,372 | 0,373 |
| *based on BEA RPP* | 0,413 | | 0,403 | | 0,372 | |

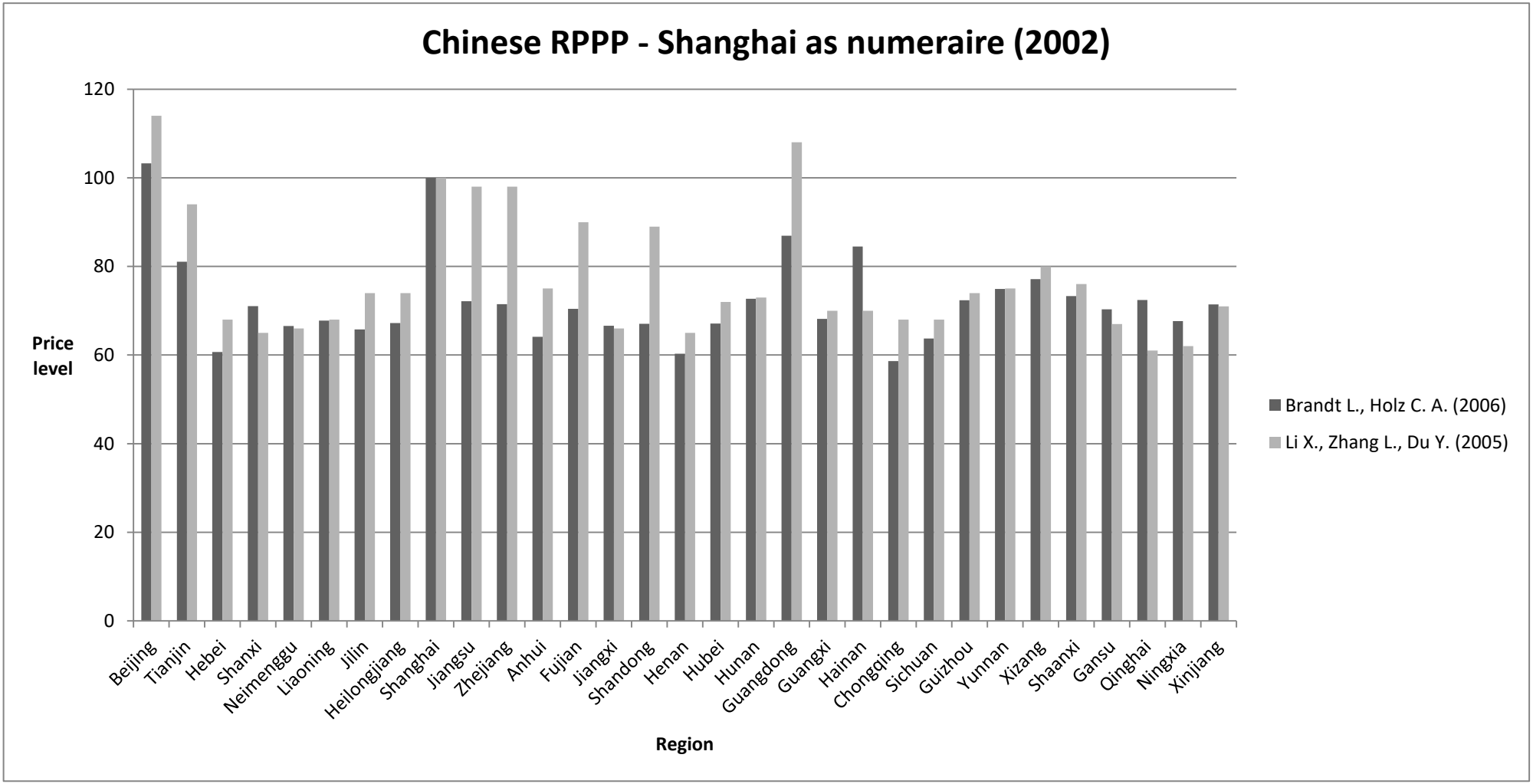
***Source:*** Author

**Appendix 6: Table 9 – Headcount ratios**

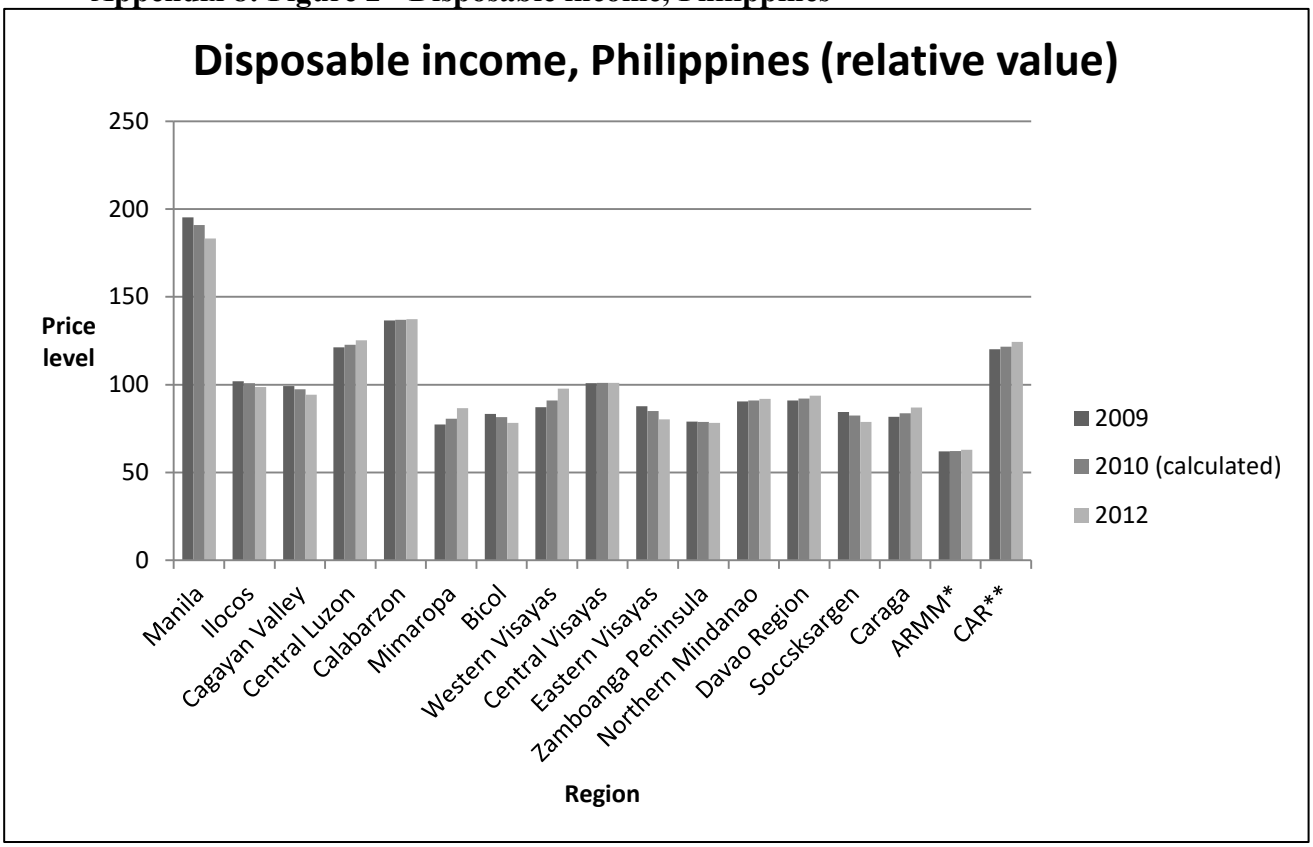
| Poverty threshold | 40% of median | | 50% of median | | 60% of median | |
|---|---|---|---|---|---|---|
| **State** | Adjusted | Unadjusted | Adjusted | Unadjusted | Adjusted | Unadjusted |
| **Australia** | 3,92 | 3,85 | 13,79 | 14,06 | 27,96 | 28,32 |
| **Austria** | 3,34 | 3,37 | 9,23 | 9,02 | 21,2 | 21,2 |
| **Canada** | 4,75 | 4,72 | 12,39 | 12,5 | 26,02 | 26,18 |
| **Colombia** | 10,48 | 10,6 | 19,35 | 20 | 34,95 | 36,1 |
| **Czech Republic** | 2,16 | 2,09 | 6,1 | 6,33 | 16,5 | 16,44 |
| **Denmark** | 2,54 | 2,53 | 6,4 | 6,32 | 19,78 | 19,72 |
| **France** | 3,82 | 3,39 | 9,99 | 9,18 | 22,47 | 21,13 |
| **Georgia** | 8,98 | 8,81 | 19,14 | 19,56 | 35,11 | 36,64 |
| **Germany** | 2,79 | 2,79 | 9,5 | 9,39 | 21,63 | 22,03 |
| **Greece** | 5,82 | 5,79 | 13,81 | 14,02 | 28,1 | 28,24 |
| **Hungary** | 4,32 | 4,61 | 10,14 | 10,58 | 22,42 | 23,56 |
| **India** | 9,39 | 9,42 | 19,41 | 19,72 | 35,41 | 35,81 |
| **Ireland** | 4,15 | 4,2 | 10,28 | 9,41 | 26,12 | 24,39 |
| **Italy** | 5,43 | 5,71 | 11,93 | 12,74 | 25,81 | 27,38 |
| **Mexico** | 10,33 | 10,53 | 19,6 | 20,18 | 33,89 | 34,32 |
| **Poland** | 3,51 | 3,56 | 9,65 | 9,62 | 22,74 | 22,69 |
| **Slovakia** | 3,64 | 3,79 | 8,15 | 8,03 | 18,7 | 18,67 |
| **Spain** | 7,57 | 7,65 | 15,31 | 15,57 | 29,21 | 29,41 |
| **Switzerland** | 3,35 | 3,3 | 9,18 | 9,15 | 22,28 | 22,07 |
| **UK** | 3,89 | 3,79 | 9,81 | 9,79 | 24,84 | 24,93 |
| *based on ONS RCPL* | 3,86 | | 9,76 | | 24,9 | |
| **USA** | 8 | 8,02 | 17,28 | 17,3 | 31,98 | 32,1 |
| *based on BEA RPP* | 8 | | 17,37 | | 31,93 | |

***Source:*** Author

**Appendix 7: Figure 1 - Chinese regional price levels**

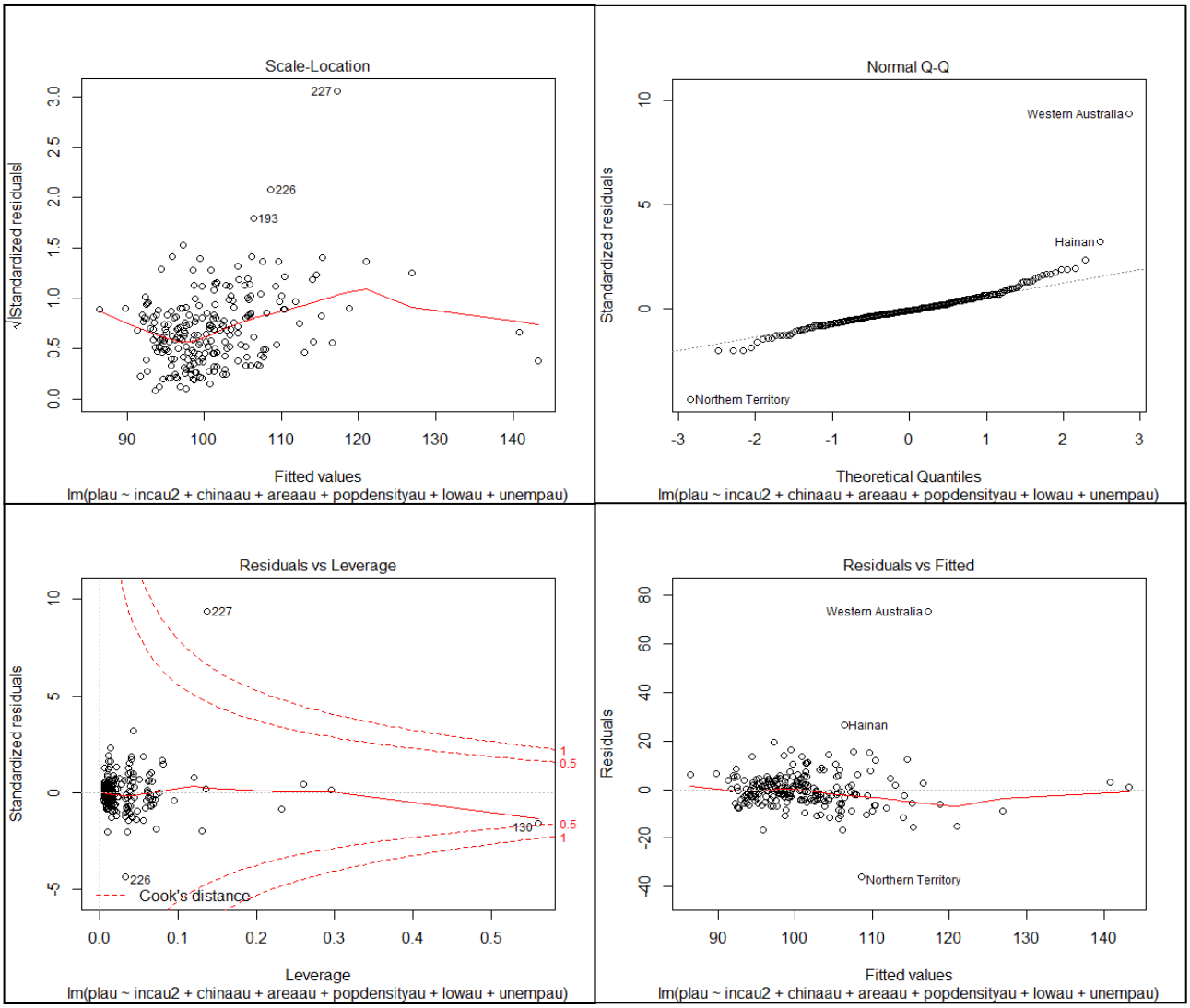


**Appendix 8: Figure 2 - Disposable income, Philippines**

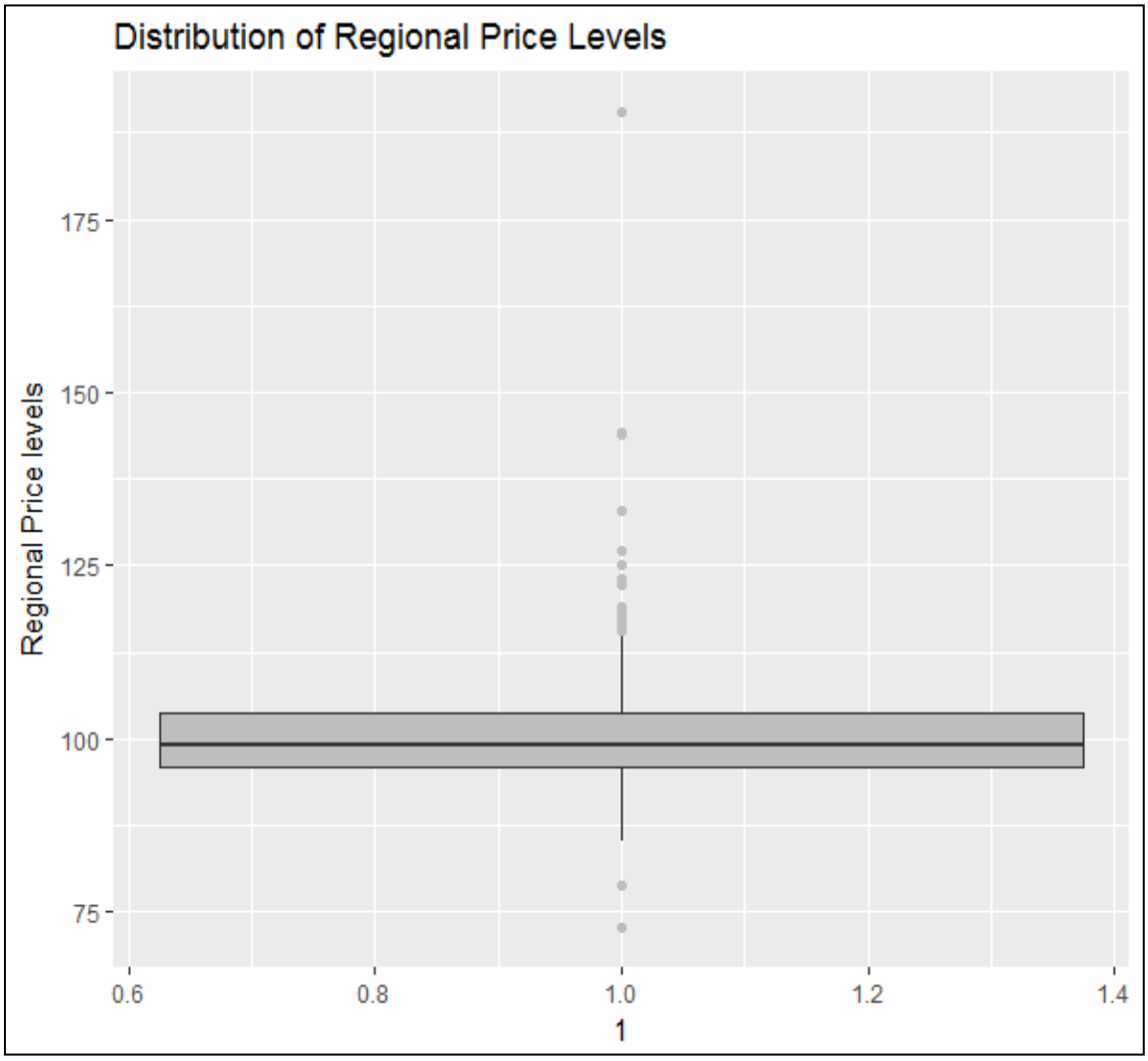



*ARMM = Autonomous region in Muslim Mindanao; **CAR = Cordillera Administrative Region

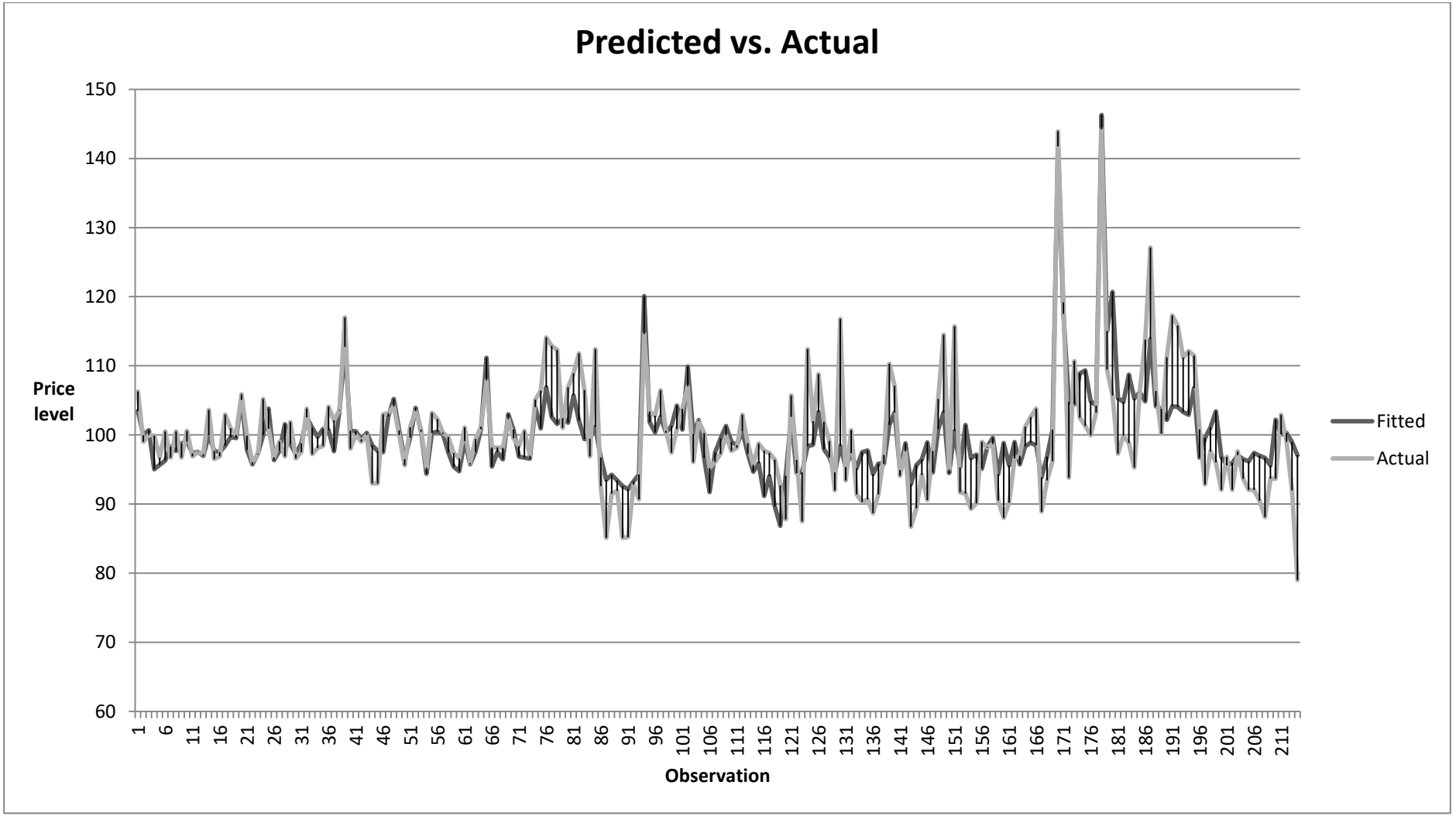
## Appendix 9: Figure 3 - Estimation results (Model 1)



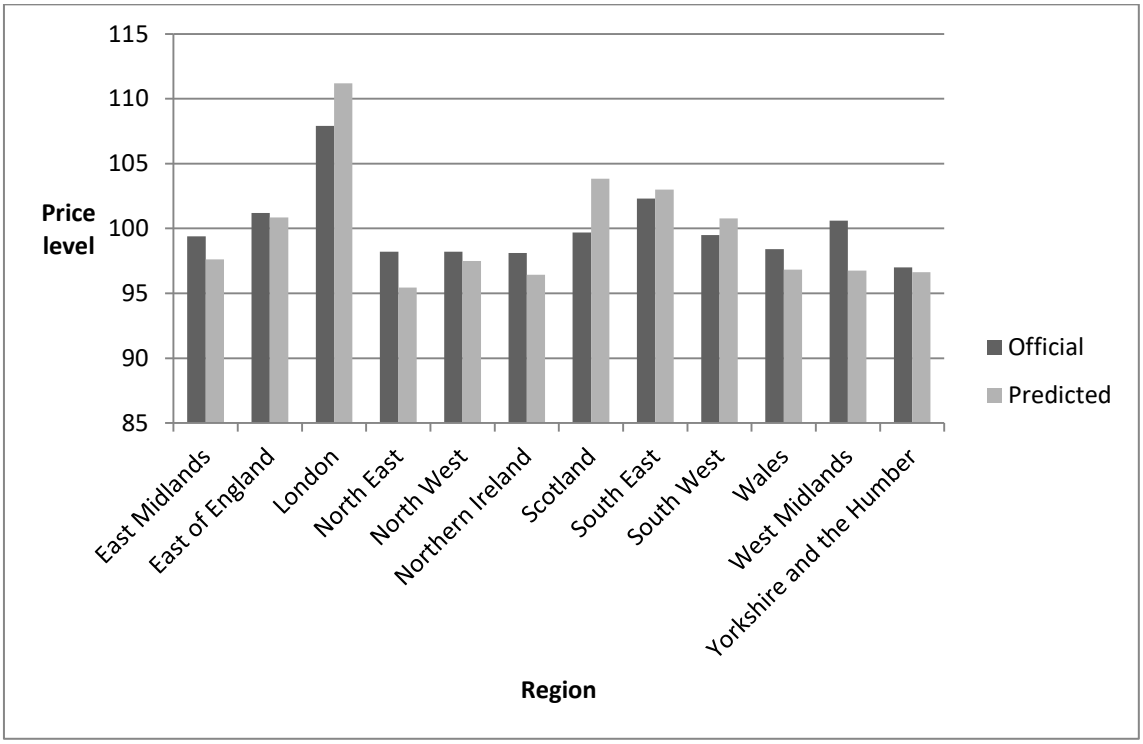
**Appendix 10: Figure 4 – Distribution of Regional Price Levels**



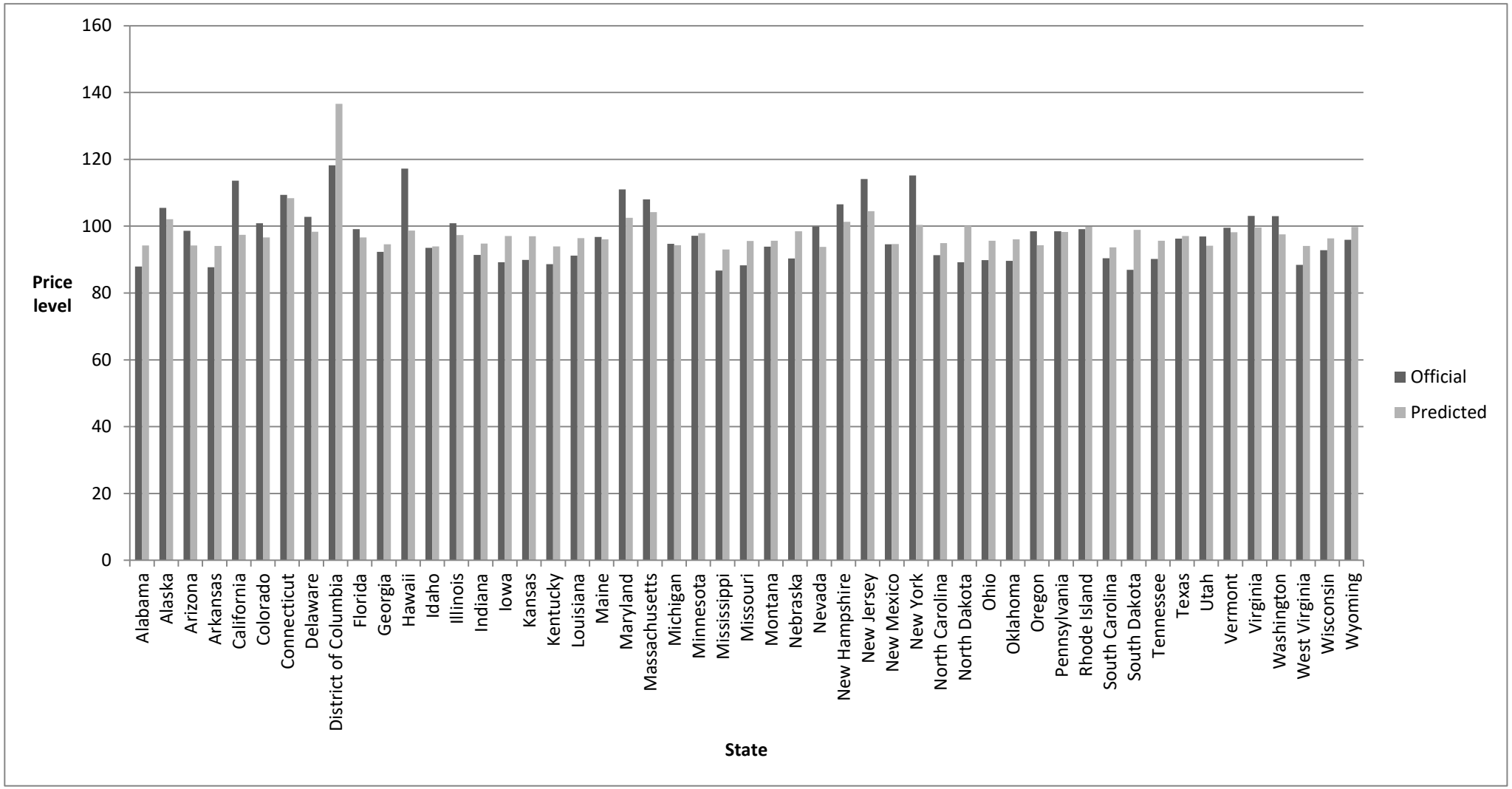
**Appendix 11: Figure 5 – Fitted vs. Actual values**



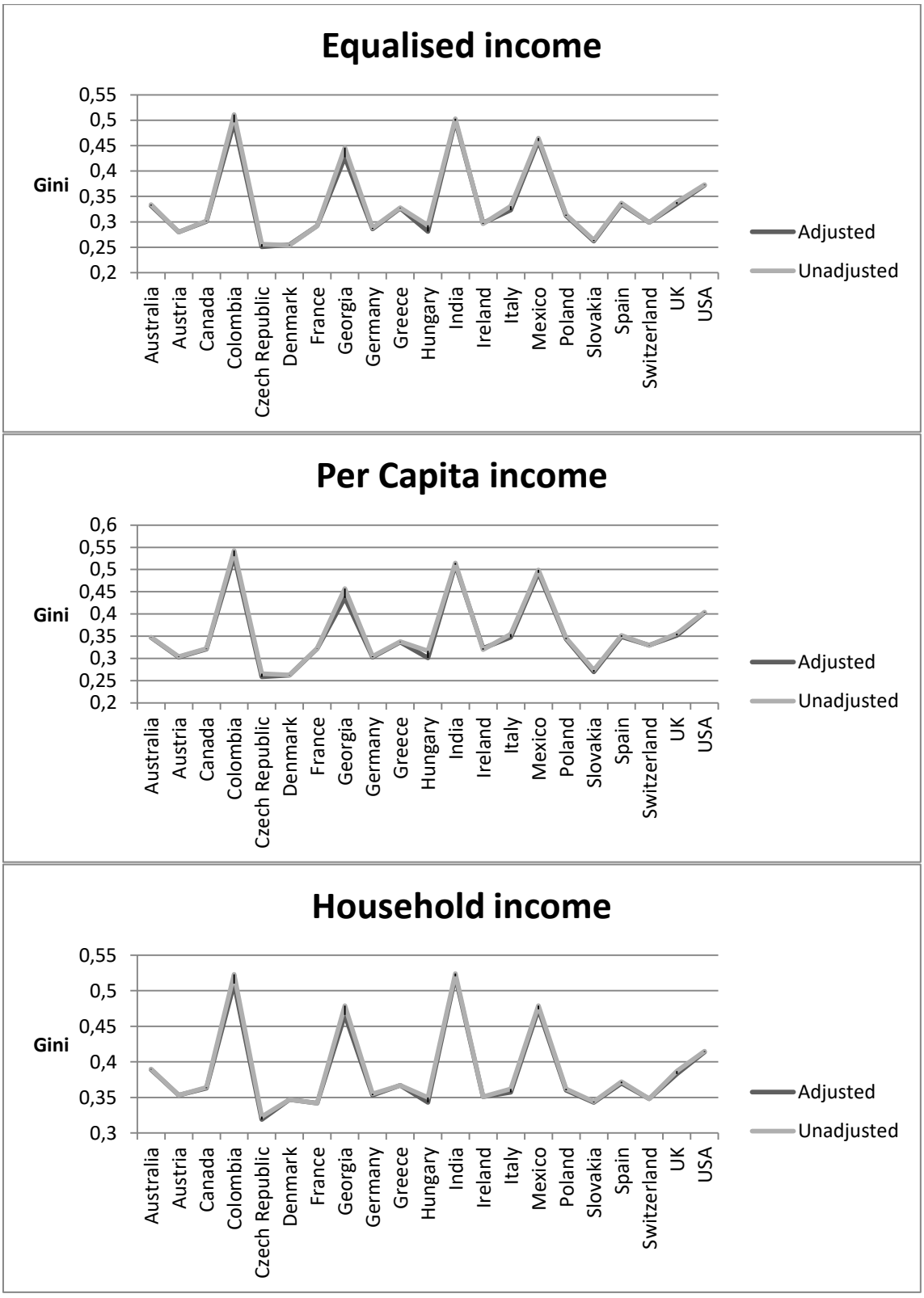
**Appendix 12: Figure 6 – Predicted vs. Official values - UK**



**Appendix 13: Figure 7 – Predicted vs. Official values – USA**



**Appendix 14: Figures 8, 9, 10 – Impacts of regional price levels on Gini coefficients**



**Appendix 15: Figure 11 - Impact of regional price levels on poverty measures (Headcount ratio)**