Charles University in Prague
Faculty of Social Sciences
Institute of International Studies

MASTER THESIS

2014 Ioana-Raluca Iordache

Charles University in Prague
Faculty of Social Sciences
Institute of International Studies

Ioana-Raluca Iordache

Health care: necessity or luxury good? A meta-regression analysis

Master thesis

Prague 2014

Author: Ioana-Raluca Iordache
Supervisor: PhDr. Jana Votapkova

Academic year: 2013/2014

Bibliographic note

Iordache, I. R. *Health care: necessity or luxury good? A meta-regression analysis.* Master thesis. Charles University in Prague, Faculty of Social Sciences, Institute of International Studies. May 2014, 84 p. Supervisor PhDr. Jana Votapkova

Typeset in LaTeX 2_ε.

Abstract

When estimating the influence income per capita exerts on health care expenditure, the research in the field offers mixed results. Studies employ different data, estimation techniques and models, which brings about the question whether these differences in research design play any part in explaining the heterogeneity of reported outcomes. By employing meta-regression analysis, the present paper analyzes 220 estimates of health spending income elasticity collected from 54 studies and finds that publication bias is of marginal concern for the literature. The model specification choices, more exactly whether a study accounts for institutional factors and advancements in medical technology, have a negative effect on reported outcomes. Moreover, the "economic research cycle hypothesis" finds support in our analysis. Lastly, the research finds that the true income elasticity of health spending is situated around unity level, which makes health care neither a luxury, nor a necessity.

Abstrakt

Tato práce analyzuje vliv národního bohatství (výše HDP per capita) na výdaje ve zdravotnictví. Existující výzkumné studie, které přináši nejednoznačné výsledky, používají různá data, metody odhadu a modely, což navozuje otázku, zda tyto metodologické rozdily hrají důležitou roli při vysvétlování různorodosti vykazovaných výsledků. Tato práce zkoumá za použiti meta-regresní analýzy 220 odhadů důchodové elasticity výdajů na zdravotnictví shromážděných v 54 studiich, a zjišťuje, že literatura na toto téma netrpí vyznamným zkreslením. Specifikace modelu, přesněji řečeno, zda studie bere v úvahu institucionální faktory a pokrok v lékařské technologii, má negativní vliv na výsledky publikované v dalších studiich. Kromě toho se analýza opírá o hypotézu "cyklu economického vyzkumu". V neposlední řadě analýza zjišťuje, že skutečná důchodová elasticita výdajů na zdravotnictví se pohybuje okolo úrovně 1, což znamená, že zdravotní péče neni ani luxusní, ani nezbytnou komoditou.

JEL Classification:	C20, I110, I180
Keywords:	meta-regression analysis, aggregate health expenditure, income elasticity
Kličová slova:	meta-regresní analýza, agregátní výdaje na zdravotnictví, důchodová elasticita

Range of thesis: 84 pages

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, May 16, 2014

Signature

Acknowledgments

I would like to thank PhDr. Jana Votapkova, my thesis supervisor, for her invaluable advice, infinite patience and support, which she provided throughout the entire process of writing my thesis. Thanks to her suggestion of using meta-regression technique in my research, I am now familiar with a method that I could further employ across many research fields. I would like to express my high appreciation for her personal involvement in my subject and for finding time in answering all my queries .

I am also grateful to PhDr. Tomáš Havránek, Ph.D. for his statistical advice and useful insights.

Institute of Economic Studies
Master Thesis Proposal

Key research question	Is health care a luxury or a necessity good? Does the literature analyzing the effect of per capita income on aggregate health care spending suffer from publication bias? Does the "economic cycle research hypothesis" apply to the studies researching the relation between per capita income and aggregate health care expenditure?
Brief description of theory	The theory of MRA is steadily grounded in statistical theory. Since MRA analyzes estimates generated by prior research, these estimates had already been previously tested and all needed assumption related to the underlying economic relationship had already been made in the primary research. The statistical theory states that the estimated regression coefficients of the primary studies, which become the MRA dependent variable, should be asymptotically normal with estimated variance equal to the squared value of its standard error. In case of publication bias, the MRA dependent variable will be correlated to its standard errors and the funnel graph of the collected estimated regression coefficients against their precision will be asymmetrical. When asymmetry is noted, then the reported estimates are heterogenous and the explanation of this heterogeneity represents the main role of multiple MRA.
Brief description of methodology	The observations of the meta-data are collected from all reported outcomes of the studies which analyze the relation between income per capita and health expenditure. Characteristics of the primary studies,i.e the estimation technique employed, the type of data used, the publication year of the paper, the explanatory variables included in their regression model, type of robustness tests performed, become the meta-independent variables, which are expected to explain the heterogeneity in reported outcomes of primary studies.
Conclusion	Publication bias is of marginal concern for the literature, while the"economics research cycle hypothesis" is confirmed. The heterogeneity in reported income elasticities of health expenditure can be explained by a number of study design characteristics. Whether a study accounts for institutional factors and progress in medical technology has a negative effect on reported outcomes. The true income elasticity of health spending is situated around unity level, which makes health care neither a luxury, nor a necessity.

Contents

1 Empirical literature on drivers of health care expenditure **15**
1.1 Trends and findings of previous research 15
1.2 Additional drivers of health care spending 17
1.3 Possible determinants of heterogeneity in outcomes 20

2 The long road to meta-regression analysis **24**
2.1 Narrative review 24
2.2 Meta-analysis 26
2.3 Meta-regression analysis 29

3 Methodology **31**
3.1 Choice of effect size 32
3.1.1 The average effect size 35
3.1.2 Publication bias 35
3.1.3 Multiple MRA 38

4 The data **40**
4.1 Construction of the dataset 40
4.2 Meta-dependent variable: the income elasticity of aggregate health care spending 42
4.3 Meta-independent variables 45

5 Results **49**
5.1 Testing for publication bias 49
5.2 Explaining heterogeneity. Multiple MRA 52
5.2.1 The "best practice" estimate 56

6 Discussion and conclusion **58**

Appendices **70**

List of Tables

4.1 Data sources 41
4.2 Average effect size 42
4.3 Meta-independent variables 47

5.1 Simple MRA results 50
5.2 Multiple MRA results 54

List of Figures

2.1 Methods of summarizing literature findings 25
2.2 Number of publications employing meta-analysis, 1986-1999 27

4.1 Frequency distribution of income elasticities 43
4.2 Funnel graph of income elasticity of health care spending 44

Acronyms

CLRM Classical linear regression model

ECO Economic Cooperation Organization

FEE Fixed effects estimator (of average effect size)

FEML Fixed effects multilevel model

GLS Generalized Least Squares

G-to-S General to specific modeling

HE Health expenditure

MRA Meta-regression analysis

OLS Ordinary Least Squares

REE Random effects estimator (of average effect size)

REML Random effects multilevel model

WLS Weighted Least Squares

"Where is the knowledge
we have lost in information?" (Eliot T.S. 1971)

" . . . only by conducting a meta-analysis can you appreciate
how messy the literature is, how inconsistent researchers are
in how they measure, analyze data . . . and just where is the
work to be done." (Hayes in Allen 2009)

Introduction

The relation between per capita income and health care costs has drawn a considerable amount of attention in the last decades. The beginning of research into this matter, marked by the study published in 1977 by Newhouse, found that the income elasticity of health care costs stands above unity, placing health care in the luxury good category. As methods of research developed and more data became available, upcoming studies, e.g. O' Connell (1996), became skeptical of the luxury good hypothesis and made claims, supported by empirical research, that health care is in fact a necessity, and its income elasticity lies below unity. Some possible influencing factors for the differences in outcomes were offered, ranging from the level of aggregation of data (regional, country-level or international), which might introduce in the analysis the problem of heterogeneity among different health care systems (Gertham 1992), the methodology employed and the number of robustness checks performed (Roberts 2000) or the independent variables included in the research model (Gerdtham & Jönsson 2000). The importance of placing health care in the appropriate category of goods lies in its social implications. If health care is a necessity, then more public involvement will be required for its equal provision to all citizens. Thus, public spending will represent a higher proportion in total health care expenditure. When health care lies in the luxury good category, it is thought that its provision would be better left to market forces and, therefore, the share of private spending in total health care expenditure would increase. Nevertheless, due to the lack of a sound theory on health expenditure determinants and the methodological complexities which developed in time, drawing a general conclusion from all existing studies becomes a difficult task.

The objectives of this thesis are manifold. First of all, the present study aims to reveal the true nature of health care, by pooling the findings of the literature on the determinants of aggregate health care expenditure. Employing meta-regression analysis, this thesis also intends to reveal whether components of a study's design influence the value of its reported income elasticity of health spending. Uncovering any such potential drivers of reported outcomes in the literature would not only help explain the heterogeneity in results, but also offer guidance for further research regarding the aspects that should be included into one's study, in order to produce a paper that comes closer to a "best practice" model. The present research also aims to see whether the

literature looking at per capita income's effect on aggregate health care costs suffers from publication bias, i.e. the preference of referees, publishers (authors) to publish (report) only results that are significant or which uncover a certain direction of the relation between income per capita and health spending. Finally, this study also tests whether the literature follows an "economics research cycle", i.e a pattern in which innovative results are preferred for an amount of time, after which empirical evidence in the opposite direction is given preference (Stanley 2008 in Havránek 2013).

The thesis is structured as follows: chapter 1 summarizes previous findings of health care spending income elasticity values, alongside differences in papers' methodology and possible factors explaining the heterogeneity in outcomes, chapter 2 offers the background and evolution of the meta-regression analysis technique, chapter 3 and chapter 4 explain the methodology and dataset employed by the present research, chapter 5 illustrates the results, while chapter 6 discusses the findings of our study and their possible implications.

Chapter 1

Empirical literature on drivers of health care expenditure

1.1 Trends and findings of previous research

Much ink has been spilled over the possible drivers of aggregate health care spending. The first one to analyze econometrically the relation between income and medical expenditure was Newhouse (1977). Using cross-sectional data from 13 developed countries, he revealed that variation in per capita income alone could explain as much as 90 % of variation in a country's per capita health spending. The income elasticity of national medical-care expenditure was found to vary from 1.15 to 1.31, making thus health care a luxury good. After this seminal work, many papers on the topic followed, being possible to classify them in distinct groups, according to the focus in research.

The initial wave of papers (Leu 1986; Parkin *et al.*, 1984; Milne & Molana 1991; Gerdtham & Jönsson 1991) focused on testing the findings of Newhouse regarding the magnitude and sign of per capita income influence on the health expenditure of a state, using cross-sectional data. These studies provided support for the conclusion reached by Newhouse: all of them found that income had a significant effect on total health care expenditure, with elasticity higher than one. Leu (1986) reported an income elasticity of 1.21, Gerdtham & Jönsson (1991) report a value of 1.43, while the results of Milne & Molana (1991) pointed to an even higher responsiveness of health expenditure to income changes of 1.74.

The next wave of research, which started at the beginning of the 1990s, focused on exploring the wider range of data sets that became available. As a result, a shift from cross-sectional to time-series and panel data sets occurred, alongside the tendency of including more explanatory variables in the regression model. Apart from income per capita, other factors considered to have an influence on total health spending were: the share of people older than 65 in the total population, the share of public expenditure in total health expenditure, technological advances, and the institutional characteristics of

the healthcare system (e.g. physicians working as gatekeepers, level of centralization of health care system, ways of remunerating physicians). In section 1.2 these additional explanatory variables will be looked at into more detail.

The use of panel data enabled the inclusion of country and time-effects in the regression model, which made previous findings highly questionable. New results pointed at income elasticity of health expenditure lower than one (Gerdtham *et al.* 1998) or very close to unity (Hitris & Posnett 1992) and the nature of health care as a luxury or necessary good started to be highly debated. Multivariate regression models displayed an even lower effect of income per capita on health expenditure, which made clear that previous research focusing solely on GDP per capita as an explanatory variable suffered from omitted variables bias. In this new context, Gerdtham *et al.*(1998) analyzed 20 OECD countries by pooling cross-sections and time-series and reached the conclusion that the income elasticity of aggregate health spending is 0.74 and thus health care is a necessity. O' Connell's (1996) results were even more radical. With a sample consisting of 21 OECD countries observed over 25 years, and using GLS estimates, he reported an income elasticity of health expenditure of merely 0.23.

The third wave of research on the determinants of health care spending focused mainly on the time-series proprieties of the data and the improvement of the methodological process (Blomqvist & Carter 1997; Karatzas 2000; Freeman 2003; Dreger & Reimers 2005; Lago-Peñas *et al.* 2013). The stationarity of the random disturbances of data series, checking for unit roots, cointegration, structural breaks and developing new, reliable robustness tests became the central point of interest in research.

One of the first papers oriented towards this new approach was Hansen and King (1996). Replicating Hitris and Posnett's study (1992), they used the same dataset over the same period of time, but checking for stationarity in the time series of income (GDP per capita) and real health care expenditure per capita. Statistical theory states that results obtained using OLS techniques may be misleading, or even entirely spurious, if the analyzed variables are not separately, or jointly, stationary. In case the data series are not stationary, the R^2 statistic reports a good fit between variables, when in fact the relation is spurious. What is more, the t and F statistics' values increase with the sample size, providing unreliable results (Phillips 1986). Hansen and King (1996) prove the variables in their model are not collectively stationary in levels, meaning that the standard OLS model employed by previous research might have generated untrustworthy estimates due to a spurious relationship.

In a similar fashion, Roberts (2000) re-examined the data and model used by Hitris (1997) and proved that his results were seriously biased due to having disregarded the time-series proprieties of the data. After applying standard unit root procedures and testing for cointegration, Roberts (2000) found non-stationarity in health expenditure, income per capita and share of total public expenditure in all analyzed countries, a finding which was robust to a variety of lag lengths. This discovery transformed Hitris'

results into a product of spurious regression problems and directed even more attention to the importance of a proper consideration of data's characteristics.

The number of tests employed for analyzing the stationarity of time series grew from one paper to another, making the methodology more complex, while the explanatory variables were again confined in most cases only to per capita income. A number of studies report the presence of unit roots and find that health expenditure and GDP per capita are first-order integrated I(1)(Blomqvist & Carter 1997; Herwartz & Theilen 2002; Freeman 2003; Atella & Marini 2006). Other studies find no unit roots in the data series for health care expenditure and GDP (McCoskey & Selden 1998) or they do identify unit roots in the data, but fail to find a cointegrating relation between the variables (eg. Hansen and King 1996). One reason for the difference in results might lie in the type of tests employed, since for instance, Hansen and King use the Augmented Dickey Fuller (ADF) test, Atella & Marini use Im et al.'s (1997) test and Blomqvist & Carter (1997) use the Phillips -Perron (PP) test. In what regards reported results of income elasticity of health care spending, a general overview of the papers which find proof of cointegration reveals that in many cases the value of this elasticity lies below unity: 0.721 (Atella & Marini 2006), 0.817(Freeman 2003 or 0. 437 (Herwartz & Theilen 2002).

1.2 Additional drivers of health care spending

Although for a long time income has been assumed to be the main determinant of health expenditure, explanations for increasing costs of health care have been searched in other parts as well.

A major determinant of health care spending mentioned by a number of studies is the share of older population in total population or the dependency rate[1].

Although in many cases the choice of model specifications is rather ad hoc, Hitris & Posnett (1992) and Gerdtham & Jönsson (1992) offer some theoretical reasons for considering the proportion of the population older than 65 years. Many papers assume older people use more health resources than other age-groups and test whether a country with a large percentage of elderly people will have higher health expenditures. Studies whose results confirm this hypothesis and find a positive influence of ageing on health spending are numerous and include Bech *et al.* 2011, Bilgel & Tran 2011, Gerdtham *et al.* 1992, Getzen 1992 and Dreger & Reimers 2005.

In contrast, a negative long-term relationship between health expenditures and ageing population was found by Samadi & Rad (2013) for a number of ECO[2] countries,

[1]The dependency rate can refer to share of population older than 65 years in total population or the sum of share of persons older than 65 years old plus the share of population younger than 15 years old in total population. It is believed that these two age-groups require more health services, and thus might incur higher health spending.

[2]ECO= Economic Cooperation Organization

with an elasticity of -0.62, in support of an optimist hypothesis where a country with elder population is considered healthy and, in consequence, people consume less costly healthcare than a country with unhealthy people. Herwartz & Theilen (2010) bring out an interesting finding, namely that the dynamic relationship between healthcare expenditure, real GDP per capita and the age structure of the population is not homogeneous across countries but relies on the age composition of the population itself. More precisely, the income elasticity is steadily larger in economies with a larger share of elder population. This proven heterogeneity among countries confirms that finding a common influence of population ageing on health care expenditure in cross-country comparisons is rather difficult.

A newer school of thought (Zweifel *et al.* 1999, Seshamani & Gray 2004, Stearns & Norton 2004, Werblow *et al.* 2007) supports the idea that what has a significant effect on health costs are the last years of life of an individual, and not the demographic age *per se*. Using Swiss data for the year 1999, Werblow *et al.* found that, on the aggregate level, age had a minor effect on an individual's health care spending, while proximity to death was strongly positively related to an individual's health expenditure. O'Neill *et al.* (2000) controlled for time to death and found no age effect on the cost of general practitioners. Furthermore, Seshamani & Gray (2004) used a longitudinal hospital dataset from Oxfordshire and by predicting costs with bootstrapped 95 % confidence intervals, found that age significantly affected hospital quarterly costs, but the magnitude was small compared to the triple growth of these costs in the last year of life. These scholars named the importance given to the effects of population ageing on increasing health expenditure a 'red herring' and an 'illusion of necessity' (Evans 1985 in Zweifel 1999) by making it appear that health care costs are inevitable and little can be done by public or private providers to stop their escalation. In addition, they bring more evidence in favor of Fries' (1980) hypothesis of 'compression of morbidity' in OECD countries, where elder generations are healthier and less infirm and use lower health expenditures than in previous decades.

A further factor included as a regressor in some models is the share of public spending in total health care expenditure. The hypothesis that has been most tested is that there exists a positive relationship between health care spending and public expenditure (Leu 1986), but once again results are mixed. Reported results of elasticities of total health expenditure to share of public expenditure range from negative values of -0.524 (Gerdtham *et al.* 1992) and -0.18 (O'Connell 1996) to positive figures of 0.018 (Lopez-Casasnovas & Saez 2007) or higher at 0.508 (Atella & Marini 2006).

Advancements in medical technology are also held accountable for influencing aggregate health care costs. While superior technology can enable some medical operations to be performed faster and more efficiently, most technology in health care focuses rather on increasing the quality of procedures or decreasing the pain of patients, which not all the time translates in increased quantity of offered health services. Furthermore, new

medical equipment tends to be greatly more expensive than previous one. Finding an adequate proxy to account for medical technology progress has proved to be a daunting task, with some scholars choosing a time index (Gerdtham & Lothgren 2000), the number of surgical procedures (Weil 1995) or the number of medical equipment. Okunade & Murthy (2002) proxy technological change in medical care by the changes in R&D spending specific to health care and identify it as the major factor on the supply-side that causes real per capita aggregate health costs to rise, with an elasticity of 0.43. Work still needs to be done on the topic, since all proxies used until now have been criticized as not being able to adequately capture the effect health care technology has on health spending.

Other explanatory factors encountered most often in the literature as possible determinants of increasing health care costs include:

- the number of physicians per 1,000 people. Some scholars believe that increasing the number of doctors will determine an increase in HCE, since physicians' services are expensive and do not compensate the gains of being healthy. For example, Murthy & Okunade (2000) find that when the number of physicians/1000 people increases by 1%, health care costs increase by 1.8%. The opposite hypothesis states that having more physicians will lead to having a healthier population that needs to use less resources e.g Gerdtham *et. al* 1998 find that when the number of physicians/ 1000 people increases by 1%, aggregate health expenditure decreases by 14%.

- share of urban population in total population. Most of the times, this variable is found to be negatively associated with health care expenditure, e.g in Wang (2009) the elasticity of HCE to urbanization degree varies from -0.49 to -0.54, while in Thornton and Rice (2008) the elasticity values are between -0.85 and -0.42. However, Samadi & Rad (2013) find on a balanced panel data of developing countries during 1995-2009 that the percentage of urbanization had a positive relationship with the health expenditures, with an elasticity of 0.742.

- education is found to exert a negative influence on health care costs. Thornton & Rice (2008) report values of elasticity ranging from -0.563 to -0.43 and Chernichovsky & Markowitz (2004) report that an increase of one in years of schooling will determine a decrease of per capita health care spending of $189,6\$$.

- female labor participation force is associated with a switch between informal and formal health care and presumed to increase aggregate HE. Pammoli *et al.* (2012) report an elasticity of health expenditure to female labor participation ratio of 0.784, significant at 1%; Bech *et al.* (2009) find a long-term elasticity of 0.0281.

- behavioral variables, i.e alcohol and tobacco consumption, obesity rates, were found to be positively associated with increase in health care costs. Employing

OLS techniques, Thornton & Rice (2008) report elasticities of 0.093 for alcohol consumption and 0.046 for tobacco consumption in a cross-section of 50 US states.

- remuneration of physicians. Remuneration of physicians by fee-for-service ways has been thought to instigate a supply-induced demand, since physicians tend to prescribe patients more services than needed. After investigating data from 19 OECD countries for the year 1987, Gerdtham *et al.*(1992) find that countries where fee-for-service systems exist incur health expenditures higher with 1.144% than in countries which do not have fee-for-service systems. Remunerating physicians by way of salaries is associated with lower health care spending.

- institutional variables. Mosca (2007) looks particularly at the effect decentralization of healthcare systems has on health care expenditure in a sample of 20 OECD countries during 1990-2000. He arrives at the conclusion that countries with a decentralized social health insurance system have had the highest health expenditures, while national health systems were allocating resources more efficiently, but not as efficient as countries with a centralized social insurance system.

Other supposed determinants of aggregate health care expenditure were tested in the literature, but will not be discussed here, since in most cases they were found insignificant. Although not directly examined by our research, some of the additional drivers of health care expenditure introduced above will be taken into account as a measure of explaining the heterogeneity in income elasticities of HE reported by the literature, as will further be seen in section 4.3.

1.3 Possible determinants of heterogeneity in outcomes

Summing up the literature on the determinants of health care expenditure, it could be said that little consensus has been reached until now. The only robust finding is the significant positive influence of per capita GDP (or national income) on per capita aggregate health expenditure. However, in what concerns the income elasticity of health spending, opinions are mixed. The same is valid for other potential factors influencing the level of health care expenditure, as could be seen in the previous section.

Various scholars researching the determinants of aggregate health spending suggested some tracks for justifying the variance in results. One specific area of concern is represented by the international comparison of health care spending, and more specifically the conversion factors used when translating the income and health expenditure levels of the countries considered into a common metric which allows comparative research. The two methods having been used until now in the literature are exchange rates and purchasing power parities (PPP) and certain scholars show differences in reported estimates when the two different conversion methods are used (Parkin *et al.* 1984). Exchange rates reflect the equalization of prices of internationally traded goods,

but cannot measure adequately the prices of non-tradable goods, such as health care. Furthermore, the exchange rate regimes alter the fraction of domestic over international prices and the measurement of prices between depreciating and appreciating countries. Using PPP over exchange rates presents some advantages, since they are calculated based on a basket of goods present in all countries. However, this does not account for the differences in the composition of the basket of goods or to the value the residents of each country attribute to its components.

Furthermore, one of the major shortcomings of the econometric research of determinants of health expenditure is the absence of a formal economic theory on the topic (Wilson 1999, p.3). As Roberts (1999) rightly points out, when results of a model are sensitive to model specification and sample composition, it calls even more for a stronger theoretical base, against which the model performance can be judged. Otherwise, variables will continue to be picked *ad hoc* and results will be as diverse as before. As Culyer (1989, p.6) puts it: "theory without history can be as misleading as evidence without theory".

A first step towards finding a proper theoretical foundation was done by Hartwig (2008, 2011), who refers to Baumol's model of unbalanced growth. This theory suggests that growth in health care expenditure levels is driven in a directly proportional way by wage increases in excess of labor productivity growth [3]. In his 2008 paper, Hartwig tests whether empirical results would be in line with this theory. He regresses the growth of per capita total current health expenditure on the difference between growth of nominal wages per employee and productivity (labeled 'Baumol variable') for a data set comprising 19 OECD countries. The regression model renders a significant value for Baumol's variable coefficient of about 1.020 and is able to explain approximately 75% of the variation of health expenditure, even after a battery of robustness checks is applied, such as including GDP per capita as an additional explanatory variable or breaking the time-series into three random sub-periods. Furthermore, in a following study, Hartwig (2011) tests another implication of Baumol's theory, this time checking if the variations in the growth rate of relative price of medical care cause variations in the health expenditure growth rate in the same direction. His findings show a positive and significant influence of the explanatory variable on health care expenditure growth, proving more support for the theory.

[3]Baumol (1967) divides the economy into two sectors, a "progressive" and a "non-progressive" one, assuming that regular growth in labor productivity, which produces new capital goods as a result of technological innovation, can emerge exclusively in the "progressive" sector. Health care, along education and other services are considered to be part of the non-progressive economic sector that produces necessities, being high-labor intensive and less physical capital-intensive. Manufacturing industries, on the other side, are part of the "progressive" sector. He also assumes that nominal wages in both sectors are connected in the long run and grow to the same rate as labor productivity in the "progressive" sector. Thus, while prices in the "progressive" sector are stable, those in the "non-progressive" sector rise so as to maintain the level of real wages in line with the productivity level. However, if the "non-progressive" sector produces necessities, such as health, for which the price elasticity is low, then a larger share of labor-force will have to move to this sector and, in consequence, a larger share of nominal GDP will have to be allocated to the sector.

Nevertheless, Hartwig's track has not been followed by more recent studies. Rather than considering Baumol's theory, latest studies (Ang 2010, Bech *et al.* 2011, Lago-Peñas *et al.* 2013, Mehrara *et al.* 2012) still use the classical variables such as income per capita, share of population older than 65 or institutional variables when trying to identify the influential factors of total health expenditure. In a certain degree, it seems understandable, since the perspective offered by Baumol's theory in the field of health care is not bright: if indeed health expenditure is bound to rise due to the "non-progressive" nature of the health care sector, then little can be done by governments and policy makers in order to stop this spending from growing.

Reported results have also been considered to vary depending on the level of aggregation of the analysis. Getzen (2000) observed that "health care is an individual good and a national luxury" and cannot be labeled strictly as one, since income elasticity seems to vary with level of analysis. After reviewing results reached in the literature, he notices that estimates using micro, i.e individual, data tend to find income elasticities of health expenditure lower than unity, and as the level of aggregation grows at macro size, so does the income elasticity of health expenditure, an observation also noticed by Pammoli *et al.* (1987). A possible explanation is that a person inside a large insured group may have little incentive to control his health expenditures, while when considered individually, the health expenditure of a household and its members depends in many cases on the available financial resources, and not on its health needs.

The structure of data, i.e. cross-sectional, time-series or panel data, has also been suspected to influence estimates. Pooled time-series allow the introduction of country-specific and time effects, which control for heterogeneity in institutional variables and technological advancements of countries being analyzed. It has been observed that, generally, research using panel data provides lower income elasticities of health expenditure than cross-sectional data does, supporting the idea that between group variation is responsible for part of the magnitude of these elasticities (Freeman 2003, p.495).

Last, but not least, the tests performed on the data and the estimation technique used might also influence results. When undergoing international comparisons, it is extremely important to tests for heteroskedasticity, since in most cases the differences in variables of the countries under observation do not respect the homogeneity assumption of the CLRM, as Murthy (1991) points out. The problem of possible outliers in samples consisting of different countries is discussed by Gerdtham & Jönsson (1992), while they show the change in results occurred by different methods of tackling the issue, i.e. eliminating outliers from the sample, creating a dummy variable accounting for outliers. Newer contribution to the topic, such as Okunade *et al.*(2004), Dormont *et al.* (2006), Barros (1998) look at health care expenditure growth rates, instead of levels.

In addition, a plethora of estimation techniques have been used when analyzing the drivers of aggregate health care spending, from simple OLS to least absolute error (LAE) in Murthy (1991), GLS in Atella & Marini (2006), WLS in Gbesemete & Gerdtham

(1992), fully modified OLS (FMOLS), dynamic OLS estimators in Freeman (2003) and many others. The same bewildering number characterizes the tests performed on the data, which makes one think these differences in methodology and estimation techniques might influence reported results in some way or another.

Among other aims, this paper will test whether some of the explanations proposed by scholars as being responsible for the vast heterogeneity in income elasticities of health expenditure results are indeed valid. By employing meta-regression analysis, the present study will examine and test systematically whether the differences in study characteristics and employed methodologies truly contribute to inconsistencies in empirical outcomes, and will generate an estimate of how much and in what precise direction these characteristics change reported elasticities. In an area flooded with studies pointing at different results, as is the literature on the determinants of health care expenditure, a meta-analysis seems to be the perfect tool to make some light.

Chapter 2

The long road to meta-regression analysis

> "Meta-analysis refers to the analysis of analyses...the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings"(Glass 1976 in Rudner *et al.* 2002).

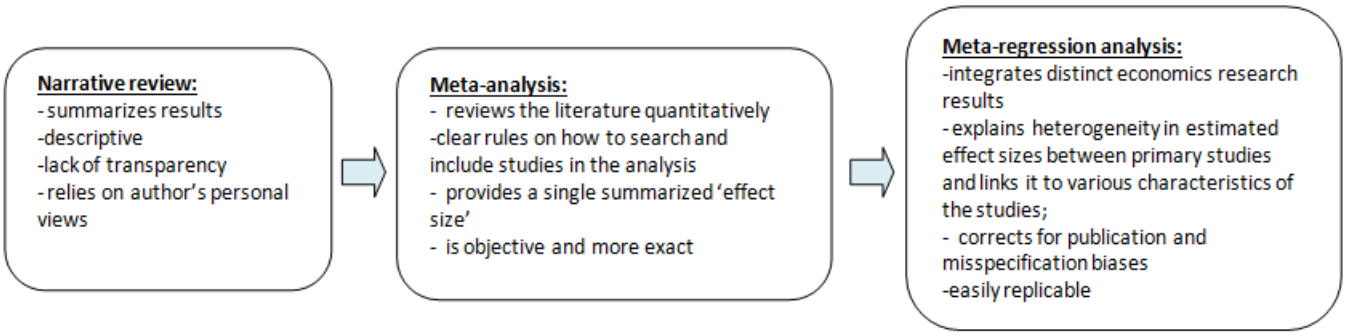
Once in a while, is useful and necessary to look back at the studies accumulated in a research field and summarize the knowledge of what has been found empirically, instead of adding a new study to the pile. By doing so, both consistencies and irregularities of previous findings can be emphasized, in order for the next study to be designed starting from what has generally been agreed upon until that moment. As Glass (1976, p.4) states in his seminal paper for meta-analysis research, there is a need for finding "the knowledge that lies untapped in completed research studies" and to use the appropriate methods "so that knowledge can be extracted from the myriad individual researches".

When trying to understand the concept of meta-regression analysis it is essential to firstly emphasize the difference and improvements this method brought to its predecessors, the meta-analysis and the narrative review. For illustrative purposes, Figure 2.1 briefly summarizes the evolution and essential differences between the methods.

2.1 Narrative review

A narrative review offers a summary of the literature on a specific field of research and compiles the results reached until a certain moment in time. Mainly, narrative reviews are descriptive and do not involve a systematic search of literature. Instead, studies are selected based on availability or the author's personal selection (Uman L.S. 2011). Typically, some papers are discarded from the analysis due to methodological inadequacies, unreliability of the data or other criteria considered important by the author, which has led many scholars to state that narrative reviews are, in most cases, characterized by

Figure 2.1: Methods of summarizing literature findings



Source:author's own compilation.

selection bias, whilst the legitimacy of choice of papers is highly questionable (Weichselbaumer 2005, p.480). A narrative review of the same literature can reach opposite conclusions, precisely due to the subjectivity and its reliance on the author's convictions or interests.

Apart from not providing comprehensive information on the methods used to select papers included in the study, narrative reviews are rather vague when it comes to disclosing information on characteristics of the data, ways in which it was collected, period of time covered by the literature or characteristics of the individual studies (Beaman 1991). What is more, a common way of reviewing the literature is by counting the number of studies supporting various results and endorsing the idea that received the most votes – the so called "vote-counting" technique. The procedure is clearly biased, since it does not take into account the research design, sample size or any other characteristics of the study's design. Last but not least, one reviewer can only deal with a small number of studies. The more information becomes available, the harder it is to evaluate it using this technique. It can thus be summarized that, while being informative and acknowledging the problems existing in the empirical economic research, conventional narrative reviews do not resolve them (Stanley 1998).

Among the encountered narrative reviews in the literature, only one focuses on analyzing the studies on drivers of health care expenditure, and that is Martín Martín *et al.* (2010) 'Review of the literature on the determinants of healthcare expenditure'. This study reviews solely papers published between 1998 and 2007, concerning exclusively OECD countries. Van Elk *et al.* (2010) also offers a short overview of the literature regarding the factors which influence health care expenditure, estimating in addition their own model. These two papers are particularly important for the present research,

since they represent the starting point for constructing the database of studies to be analyzed by our MRA. More details regarding the dataset will be offered in chapter 4.

2.2 Meta-analysis

Meta-analysis, on the other hand, is "a more rigorous alternative to the casual, narrative discussions of research" (Phillips & Goss 1995, p. 322). While, as the narrative review, it analyzes and synthesizes the results of the literature on a certain topic, it does so using statistical tools. Meta-analysis does not replace the narrative review technique, but it rather accompanies and grounds the results reached by it. It is more objective and quantifies the results reached in the literature, providing a measured averaged of each study's outcome.

Although the statistical basis of meta-analysis was built well before his time [1], the statistician Gene V. Glass is recognized as the one who coined the term"meta-analysis". He defined it as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass 1976). In this seminal work, he calls for attention being drawn on the importance of extracting knowledge from gathered studies and although his investigation was focused on the analysis of the educational area, meta-analysis soon became adopted in many fields of research. Due to the improvements it brought to traditional narrative reviews, starting from mid 1980's the number of papers using meta-analysis grew exponentially (Figure 2.2).

The central element in meta-analysis is the "effect size", which could show the result of an intervention, the power of a relation between two variables or the estimate of a distinct value (Borenstein 2011, p.4). As defined by Glass, the effect size represents the mean difference on the outcome variable between treatment and control groups divided by the control group's standard deviation:

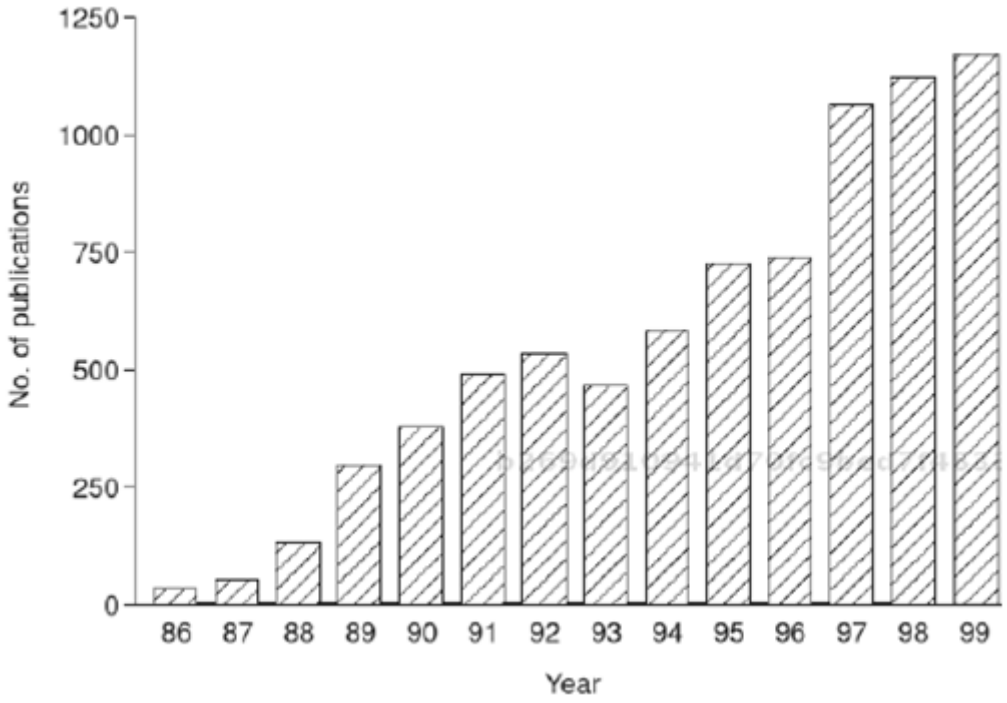
$$g = (\mu_e - \mu_c)/\sigma \tag{2.1}$$

where μ_e is the mean of the experimental group, μ_c represents the mean of the treatment group and σ is the standard deviation of the control group (Glass 1976, p.6).

The effect size represents the common denominator of the literature being analyzed and is used to compare estimates from different studies. Research might report estimates measured in different metrics, i.e elasticities, semi-elasticities, correlations, regression coefficients, but they can all be transformed into the common metric which is the effect size and later be compared. This idea is fundamental in understanding how meta-analysis, and further meta-regression analysis works, since the effect size represents the dependent variable of the meta-regression model.

[1] The first methods of quantitatively summarizing the results reached by all studies in a certain field of research was first introduced by George Biddell Airy in his 1861 book 'On the algebraical and numerical theory of errors of observations and the combination of observations'

Figure 2.2: Number of publications employing meta-analysis, 1986-1999



Source:Egger et al., 2008, p.23

Basically, each study in the literature considered for evaluation is characterized by an effect size. After weighting each study according to its effect size's precision, namely the inverse of its standard error, an average effect can be computed, which is a more precise and robust estimate of the "true effect" than the individual estimates provided by each primary study. This is, in fact, the major aim of meta-analysis: providing an estimate with an increased statistical power that could reveal statistical significance in a research field where conclusions are mixed and vague.

For a long time, meta-analysis was used in the field of medical research and psychology, where performing new trials was a rather expensive process, both in money and lives (Stanley 2001). Borenstein (2009) emphasizes how pharmaceutical companies summarize the data from studies on the efficacy of a drug, providing a more powerful and precise estimate of its effect. Meta-analysis proves helpful also when evaluating the side effects of a drug, especially since cases are more rare, and thus primary studies are more scarce (Borenstein *et al.* 2009).

Meta-analysis can solve a number of shortfalls present in narrative reviews:

- the lack of transparency and high subjectivity problem. A meta-analysis will offer all the details regarding the techniques used to search the literature and

criteria on which studies are included or excluded from the analysis, the methods used for performing statistical analysis and the mechanism for disseminating the results (Borrenstein 2009, preface). Also, the author's own views do not have such an influence on the results as they do in narrative reviews, since the weight, corresponding to the importance, assigned to each study is based on mathematical criteria;

- the replication problem. Since meta-analysis offers detailed instructions on the methodology employed, research can be replicated by scholars interested in studying the same topic and observing that results produce various changes in the data set, variables chosen or methods used;

- the aggregation problem. The more studies included in a review of the literature, the harder it becomes to analyze them through descriptive methods. The human mind could generate an overall conclusion and summarize subjectively the results of 30 studies, but when faced with the analysis of 200 studies or more, it is faced with an impossible situation (Rudner*et al.* 2002). Meta-analysis makes things possible, employing statistical methods and statistical software, coding the results of each study and using a common metric for all outcomes, in order to make them more comparable;

- the ambiguity problem. Very often, narrative reviews provide vague results, that are difficult to use by policy makers. By providing a measure of the average outcomes, the results of a meta-analysis are more usable.

However, meta-analysis has its limitations. One often-mentioned criticism refers to the inadequacy of drawing conclusions by comparing and pooling studies that use different measurement techniques and variables. Also, meta-analysis often considers multiple outcomes that come from the same study, which might over-represent some papers and under-represent others. Another problem pointed out is that meta-analysis includes both unbiased, robust studies and papers whose design is questionable. This aspect of methodology is suspected of providing an overall estimate that might be biased and not representative. Finally, the issue of publication bias arises, referring to the fact that meta-analysis results might be biased also due to the fact that published research might favors studies reporting a certain result, while those papers which do not follow the pattern seldom get published (Phillips and Goss 1995).

The concepts of treatment and control group, as were introduced by Glass in defining the effect size (Equation 2.1), might seem unsuitable for economic research, as in the present case. Fortunately, meta-analysis started being used on a large scale in social sciences as well, of special importance being Stanley and Jarrell's initiative (1989) of employing meta-analysis in the field of economics. Trying to address and offer solutions to some of the limitations of meta-analysis listed before, they developed the specific procedure of assessing empirical economic research: meta-regression analysis.

2.3 Meta-regression analysis

The first article on the use of meta-regression analysis in the field of economics was published in 1989 in the Journal of Economic Surveys. Unlike meta-analysis, which only offers a weighted average of the studies' effect size, MRA goes further and tries to clarify and quantify the causes of heterogeneity in estimated effect sizes between the primary studies. MRA works as a typical regression analysis, in which the population is no longer composed of individuals or countries, but of primary studies. Different estimates of a given coefficient inside a primary study may be treated as individual observations(Rose & Stanley 2005).

In the last decade, MRA was employed to look at a variety of topics: the effect of state and taxes on economic development (Phillips & Goss 1995), the factors that explain local privatization (Bel & Fageda 2007), the effect of common currency on international trade (Rose & Stanely 2005), gender wage differences (Weichselbaumer & Winter-Ebmer 2005), the effect of immigration on labour markets in the host country(Longhi, *et al.* 2005), the link between economic freedom and economic growth (Doucouliagos 2005) or the determinants of horizontal and vertical spillovers from FDI (Havránek 2013).

By definition, MRA includes all studies, published and unpublished, on a specified topic (Stanley & Jarrell 1989). The differences among primary studies, e.g. the methodology employed, the ways in which variables are measured, robustness of estimates or tests applied, should not represent a reason to eliminate studies from the data set, but instead offer more grounds to include them and control for their effect on the meta-dependent variable – the effect size. If papers are excluded from the analysis, the reasons of the choice have to be explained in detail (Stanley 2001). This procedure greatly minimizes the subjectivity of the selection process and promotes replication by others interested in the meta-analyzed area of research or by the method itself.

In an MRA, a central point of interest is held by the meta-explanatory variables. In the end, a researcher undergoing a meta-regression research is interested in knowing how different model specifications or characteristics of the primary studies influence their estimates. Some possible characteristics of a study that could be employed as meta-regressors are: a researcher's age, gender, nationality or level of income, data used, variable definition, year of publication of study, source of funding of research, methodological characteristics of the paper etc. (Stanley *et al* 2006, p. 284).

Apart from being able to explain heterogeneity in reported results from a specific field of research, MRA addresses another important problem: publication bias[2]. When aggregating the results of various papers, the main bias the researcher is faced with is the fact that literature might be missing those studies that do not provide the results editors or researchers favor. Type I bias refers to the case when papers reach results bearing a different sign than the one preferred by editors or authors, e.g. if a study

[2]Staneley & Doucouliagos (2012) also refer to the term 'selective reporting bias', since authors themselves choose not to report insignificant or not preferred results

would reach the conclusion that per income capita has a negative influence on per capita health expenditure, there might be a tendency to discard these results, since the consensus among researchers is that an increase in income levels triggers an increase in per capita total health care expenditure. Type II bias appears when papers reporting insignificant results do not get published, since editors find they are not a 'good story' for publication (Havránek 2013, p.15).

Publication bias gives overall reported estimates a distorted value, which can have considerable consequences for policy makers. If a literature is affected by publication bias, then the MRA sample will be biased and the probability of committing a statistical type II error, i.e accepting a false null hypothesis, is higher. MRA can correct for publication bias and provide an estimate of the true effect beyond the contamination of publication bias. More about the ways in which this can be achieved will be presented in the next chapter.

To sum up the main points, it could be stated that narrative reviews are a good tool for getting a general overview of the results reached in a certain field of research, but the examination is excessively subjective and once the number of evaluated papers grows, it becomes difficult to offer reliable outcomes . Meta-analysis manages to quantify the results reached in the literature and offers an average effect size, but it is highly biased, since it does not account for the results of those papers which might not have been published. While it offers a quantitative measure of the outcomes from a research area, it does not try to account for the primary study characteristics that might influence reported estimates. Meta-regression analysis seems to be the best alternative when trying to summarize and interpret the findings of economic research. Its main shortcoming, the exposure to publication bias, is also encountered in narrative reviews or simple meta-analysis, but unlike these approaches, MRA uses objective, replicable methods, identifies and removes publication bias and explains the heterogeneity in research findings (Stanley & Doucouliagos 2008).

Chapter 3

Methodology

As previously mentioned, the central idea in MRA is the effect size. This represents the estimate of the relation of interest, which in our case is the influence income has on aggregate health expenditure levels. Primary studies might report estimates using different measures, e.g. elasticities, semi-elasticities, partial correlations. These should be transformed into a common metric to allow further comparisons and assessments of overall results. Once the conversion completed, that metric will represent the effect size and will be introduced in the MRA model as the dependent variable.

The literature assessing the drivers of health care expenditure generally uses a variation of the following regression model:

$$ln\text{PHE}_{it} = \beta_0 + \beta_1 ln\text{PCI}_{it} + \beta_2 \text{Demand}_{it} + \beta_3 \text{Supply}_{it} + \sum_t \text{Y}_t + \sum_i \text{P}_i + u_{it} \quad (3.1)$$

where PHE_{it} is per capita health expenditure, PCI_{it} is per capita gross domestic product, Demand_{it} represent the demand side variables (e.g. percentage of population over 65 years, mortality rates, life expectancy, urbanization), Supply_{it} represent the supply side variables (e.g. number of physicians per 1000 people, no. of acute care beds per 1000 people, relative price of health care), Y_t are year fixed effects, P_i are country fixed effects, u_{it} is the error term and i and t index the cross-section and time period, respectively.

The reader will be asked to interpret this model only as the frame upon which are constructed the regression equations encountered in the literature. Depending on the type of data used and independent variables considered by researchers, there exist variations. For example, many studies focus solely on the influence of income on health spending and as a result the only independent variable included on the right-hand side is PCI (eventually, alongside time and country effects for panel data). The choice in additional supply and demand-side variables also differs, in accordance to what the researcher might consider to be a possible influence on health spending. Finally, papers employing cross-sections can not include any country or time-fixed effects.

3.1 Choice of effect size

What we are interested in from Equation 3.1 is the estimate of β_1, which shows the effect per capita income exerts on aggregate health care spending. It has become rather a consensus in the literature of the drivers of health spending that the log-log functional form is preferred to others, which means the coefficients are the direct estimates of elasticities. Since elasticities are non-dimensional, it is possible to compare them within and between different studies, and thus prove to be an adequate effect size for our MRA. The elasticity exhibits the two important proprieties of effect size, namely it measures the effect of one variable on another, and is comparable between and within studies, since is a dimensionless value (Stanley & Jarrell, 1989). Thus, the estimated health spending income elasticity reported by each primary study was collected and introduced in our MRA analysis as the dependent variable.

Where studies use the log-log functional form, estimated coefficients of the income variable could directly be interpreted as elasticities. However, a number of papers employ models in levels. In some cases, alongside the estimate of coefficients, they also offered information on elasticities values (e.g. Parkin *et al.* 1987). In case elasticities were not reported, it would be possible to calculate their value at the mean[1]. However, the studies we encountered did not normally report any information on sample means of independent and dependent variables, which meant that we could not include these papers in our dataset, as the effect sizes would not be comparable.

The next issue appears when referring to standard errors. In MRA, the standard errors are needed in order to appropriately weigh each primary study when computing the averaged effect size or when estimating the multiple MRA. If a primary study uses the double-log function and reports the standard errors of its coefficients, these can further be employed directly as the standard error of our effect size. In some cases, papers did not report standard errors, but offered t-values instead. Standard errors were then computed using the formula:

$$t = \eta_1 / SE_{\eta_1} \tag{3.3}$$

where η_1 represents the estimated income elasticity, t is its t-value and SE_{η_1} stands for its standard error.

Other studies only report the p-values of the estimates. When degrees of freedom are mentioned or can be computed from the information provided by the paper, we derived the t-statistics using the TINV function in Excel, and then computed the standard error. We also happened to encounter studies that only exhibit the level of significance, using

[1]The elasticity can be calculated using the formula

$$\eta_1 = \alpha_1 \overline{X} / \overline{Y} \tag{3.2}$$

where $\overline{X}$ represents the sample mean of the explanatory variable and $\overline{Y}$ stands for the sample mean of the dependent variable.

***,**, and * for significance at 1 percent, 5 percent and 10 percent levels. In this case, we followed the example of Greenberg *et al.* (2003), and assumed the p-value is located at the midpoint of statistical significance scale. For example, an estimate significant at 1 percent level was assumed to have a p-value of 0.005; an estimate significant at 5 percent level was given a p-value of 0.03, while an estimate significant at 10 percent level was assumed to have a p-value of 0.075. This method does introduce some measurement error in the dataset, but it was considered to be a better option that eliminating these studies altogether.

Moreover, those studies not measured in log-log functional form, yet still reporting elasticity values, did not usually declare the standard error of the elasticity, but only the regression coefficient's one. Regarding this aspect, Stanley & Doucoulagios (2012 p.27, further referred to as S & D 2012) recommend that in case the standard error of elasticity cannot be computed [2], due to unreported needed information, it is possible to use the standard error of the coefficient instead. They claim that the precision error introduced by using this method would not be as a big concern as the publication or misspecification biases the literature might suffer from. When checking for publication bias and heterogeneity in reported income elasticities, we considered the t-values of the regressions coefficients, from where it was further easy to compute the standard error of the elasticity, using eq. 3.3.

When a paper reports both long-run and short-run income elasticities of aggregate health care costs, we collect only the long-run estimates, again following S & D's (2012) recommendations. We also find it to be a more appropriate estimate, since the short-run elasticities measure the dynamic relation between income and health care expenditure and the rate at which the relation between the two would reach its equilibrium state after an alteration in one of the variables. The long-run elasticities, on the other side, estimate a more stable relation between the two variables and can suit our analysis better.

Finally, the issue of multiple estimates reported in a primary study should be addressed. Numerous studies included in our dataset report multiple estimates, either resulted from the estimation of nested models, usage of different techniques or employment of different data. S & D (2012) present various possibilities of tackling this issue. The first one would be to report one estimates per study, coming from the model explicitly preferred by authors themselves or resulting from the most robust and with the highest explanatory power model. This approach will generate a scarcer data set, and might also introduce selection bias, since the meta-analyst must select an estimate based on certain judgments.

The second possibility would be to collect all effect sizes reported in a study and compute the weighted average of them, alongside the weighted average of the standard

[2] This can be achieved using the formula $var\eta_1 = \frac{\overline{X}^2}{\overline{Y}^2} var\alpha_1 + \alpha_1^2 \frac{\overline{X}^2}{\overline{Y}^2} var\overline{Y}$, where $var\eta_1$ = the variance of elasticitiy, and $\overline{X}$, $\overline{Y}$ are the sample means of the independent and dependent variable, respectively.

errors. The magnitude of the effect sizes, in this case, will be different than when using the first approach, which will clearly provide different results, closer to the overall value of all estimates reported in the literature. Nonetheless, this approach does not increase the number of observations of the MRA and also disregards the differences in methods used to estimate the reported effect sizes.

The third method, which was the one employed in the present study as well, consists of considering all estimates reported by each study, coding for the differences in methods and estimation techniques which were used to obtain them. One advantage is that it greatly increases the size of the database, offering thus more degrees of freedom and space for testing the effect of more meta-independent variables on the effect size. Secondly, since differences between methods used for computing each effect size will be controlled for, it also has a greater potential to explain the variability in estimates of health spending income elasticities and identify those meta-regressors responsible for the heterogeneity among reported estimates of the elasticity.

When MRA started being used, it was believed that the best option was to pick one observation per study(Stanley & Jarrell 1998). Nowadays, conventional practice agrees that the use of all estimates, which are weighted inside a study, while also modeling for within-study dependence by using multi-level MRA, is the best practice. Including more than one estimate from a single study brings about the possibility of dependence in the data, but MRA has a solution for this problem.

In meta-regression analysis, the studies included are assumed to be a random sample of the distribution of effect sizes. The mean of this distribution is represented by the combined effect estimates (Borrenstein *et al.* 2007, p.4). Some critics of meta-regression analysis claim that combining and averaging the results of studies which differ in quality might provide biased estimates. This issue could easily be fixed by introducing in the MRA a measure of quality of the estimates collected from the literature. For our research we use the precision, i.e. the inverse of and estimate's standard error, as such a measure. This way, studies which report estimates of lower quality, i.e with a higher standard error, will be given less weight when computing the overall effect estimated in the literature.

The weighted average effect size, which in this case is the average of all health spending income elasticities reported in the literature, has been computed using the formula:

$$\eta_w = \frac{\sum w_i \eta_i}{\sum w_i} \tag{3.4}$$

where w_i are the weights used and η_i are the estimated elasticities in each primary study. When it comes to weights, it was suggested that the optimal case would be to use the inverse of the estimates' variances (Stanley & Doucouliagos 2012, p.46; Borrenstein *et al.* 2007, p.7). As a sensitivity analysis, the average effect size was also computed using as weight the inverse of the number of effect sizes a study reports. For example, if

a paper provided 10 estimates of income elasticity on health expenditure, each estimate was weighted by a measure of 1/10. Next, the standard error of the average effect was computed as the square root of the reciprocal of the sum of the optimal weights (S & D, 2012, p. 46), which is:

$$\text{S.E}_{\eta_w} = \sqrt{1/\sum w_i} \tag{3.5}$$

3.1.1 The average effect size

The computation of the average effect size was done in two ways: using fixed effects average estimators (FEE) and random effects average estimators (REE). It must be noted that in MRA fixed effects and random effects estimators are not the same as the ones encountered in panel data regression analysis. We will use the FEE and REE acronyms when referring strictly to the averaged effect estimator method.

FEE assume that all effect sizes come from the same "true effect size" and have been extracted from the same population. When computing the average weight, each effect size is weighted by its squared precision. Thus, an estimated fixed effect will equal the true effect size plus a within study error ($T_i = \Theta_i + \varepsilon_i$). In contrast, REE weigh each study by the construction ($SE_i^2 + S_h^2$), where SE_i^2 estimates the within-study variance and S_h^2 estimates the between-study variance. Random effect estimators assume that the true effect size varies from study to study. When using REE, one estimates the mean of a distribution of true effects, since there is no unique true effect for all studies. That is why the weights under the REE have to account for both within and between-study variances ($T_i = \Theta_i + \zeta_i + \varepsilon_i$)(Borrenstein *et al.* 2007, p.5). The average effect size computed with the REE still uses Equation 3.5, but this time the total variance will also account for between study variance. This will become:

$$\text{variance} = \sum w_i \eta_i^2 - \frac{(\sum w_i \eta_i)^2}{\sum w_i} \tag{3.6}$$

The choice between FEE and REE should be made according to the studies included in the database. If one believes the studies have the same underlying true effect, FEE could be used. In comparison, if data is collected from studies performed by others, chances are the data is not identical and the true effect size differs from study to study. In this case, REE would be more appropriate. Since our database is composed of studies which use data from different regions and different methodologies, we have reason to suspect the underlying effect size differs, which is why REE will be a more appropriate way of averaging the income elasticities of health spending reported by the literature.

3.1.2 Publication bias

One of the important contributions of MRA to normal meta-analysis is the sorting of publication bias and modeling of heterogeneity. Publication selection appears when

studies are left out of the public reach, because they are not published. Two reasons can lie underneath this bias: either the results have a direction which is not preferred by referees, editors or authors, either the outcomes of the study are not significant and therefore do not represent a "good story" for publication (Havránek 2013, p.15).

As a first step in uncovering publication bias, we graphically represented our data using a funnel plot. Income elasticities were plotted against their precision, i.e. inverse of standard error. Theory states that if the scatter plot resembles an inverted funnel shape, with more precise estimates closer to the average value, and it displays a symmetrical distribution of effect sizes, then most probably the literature does not suffer from publication bias [3]. Otherwise, there is a high risk the literature is contaminated by publication selection (S &D 2012).

Apart from the funnel graph, which only offers a glance at publication bias, MRA can test for its presence in a more formal way, employing a simple regression which reformulates the funnel plot. This bivariate regression emphasizes the relation between the estimate of income elasticity and its standard error. In the absence of publication bias, reported effect sizes will be independent of their standard errors, as is assured by random sampling theory. In the case of publication selection, the estimated effect size of each primary study will be correlated with its standard error, *ceteris paribus* (S & D 2012). The explanation for this correlations, as given by Stanley (2012), is that researchers who have small samples and thus large standard errors will have to look more for larger estimates and test with various model specifications and statistical techniques in order for their outcomes to have a statistical significance. On the other hand, researchers with more numerous samples and smaller standard errors have practically infinite model specifications to find statistical significance and do not need to search so hard (Stanely 2008). Also, since journals are more likely to publish significant results, the standard error of an estimate is normally considered a good indicator of precision or publication bias. The idea can be translated into:

$$\text{effect}_i = \beta_0 + \beta_1 \text{SE}_i + \varepsilon_i \tag{3.7}$$

where effect_i is the effect size, i.e. the income elasticity estimate reported by study i, SE_i is its standard error and ε_i is the error term. The term $\beta_1 SE_i$ models for publication bias, while estimates of the intercept β_0 provide corrections for publication selection (S& D, 2012, p.61). The corrected estimate will basically represent the difference between the unweighted average effect and the estimate for publication bias ($\beta_1 \text{SE}_i$). Previous scholars have found that in fact, when correcting for publication bias, a better measure is provided when using the variance instead of the standard error of the effect sizes and we will estimate our model by using also this variation.

Since in equation Equation 3.7 the dependent variables is the effect size and the

[3] Publication bias will still have to be tested by means of regression analysis.

independent variable is a sample of estimates of its standard error, the error term ε_i will not be similarly and independently distributed. In consequence, the model is heteroskedastic. In order to eliminate this problem, we estimated its WLS version, in which all components are weighted by the inverse of the squared standard error of the corresponding elasticity.

Attention was paid to the issue of dependence among estimates of elasticities, to which we briefly referred when computing the FEE and REE average effect sizes. Most of the studies included in our database report more than one estimate of income elasticity of health expenditure, which raises questions regarding the dependence between these studies, which might make Equation 3.7 misspecified. As a robustness check, the simple WLS MRA was also estimated by ways of a mixed-effects multilevel model, which controls for unobserved between-papers heterogeneity (S& D 2012 p. 69). This model is an equivalent of the random-effects model used in panel data regressions. The formula we employed for our estimation is:

$$\text{effect}_i = \beta_0 + \beta_1 \text{SE}_{is} + \mu_s + \varepsilon_{is} \tag{3.8}$$

where i stands for the estimate number,s is the study number, μ_s is the unobserved study effect and ε_{is} represents the between-studies unobserved effects.

The mixed-effects multilevel model is appropriate for MRA, as it considers unbalanced data and thus employs the restricted maximum likelihood estimator (MLE) instead of generalized least squares (GLS) (Nelson & Kennedy 2009 in Havranek & Irsova 2010, p. 10). A test on the null hypothesis: $H_0 : \beta_1 = 0$ will reveal whether the literature on drivers of aggregate health expenditure suffers from publication bias (if the null hypothesis is rejected). This test is called the funnel-asymmetry test (FAT) and its results are presented in section 5.1.

Once the FAT test revealed the presence of publication bias, although a low one, we continued our analysis by checking whether there exists a genuine empirical effect, beyond the small distortions in estimated effect sizes brought by publication bias. This was achieved using the precision-effect test (PET), which tests the null hypothesis : $H_0 : \beta_0 = 0$ of the same simple MRA equation (3.7). Rejecting the null hypothesis ensures there is a true empirical effect of income on health care expenditure. It should be noted that the FAT and PET test have their shortcomings; more precisely, FAT is known to have a low power (Stanley 2008), while the PET can point at the existence of a genuine effect size, even when there is none (type I error).

Finally, after proving the existence of a true effect size, beyond any influence of publication bias, we estimated its value using the precision-effect estimate with standard error (PEESE) test. It was mentioned before that various simulations performed by scholars found that using as independent variable in (3.7) the variation of the estimated effect size, and not simply its standard error, provides more reliable results for the

coefficient on precision if publication bias is proved to be even slightly present(Stanley & Doucouliagos, 2011). The equation used for estimating the model is:

$$\text{effect}_i = \beta_0 + \beta_1 \text{SE}_i^2 + \varepsilon_i \tag{3.9}$$

The PEESE estimate should be used in those cases where there is a suspicion that a true underlying effect size exists. This means that if the PET test rejects the null hypothesis, then one can estimate the PEESE and it will provide better estimates of the genuine effect than PET would. In our case, since the PET test provided results which indicated the existence of a true effect size, we were able to estimate its value more reliably by using the PEESE test.

3.1.3 Multiple MRA

Once publication bias has been identified and a measure of the true effect has been estimated, a further step in analysis was to uncover what causes the excess heterogeneity in reported estimates of income elasticities of health care expenditure. For this purpose, different characteristics of the primary studies believed to affect the reported outcomes were collected and coded and then employed in a multivariate MRA.

The multivariate MRA model is similar to the simple MRA one , with the difference that it includes the extra-explanatory variables:

$$\text{effect}_i = \beta_0 + \beta_1 \text{SE}_i + \sum \beta_k Z_{ki} + \varepsilon i \tag{3.10}$$

where effect_i represents the estimated income elasticity of health spending, β_0 is the intercept and estimates the true income elasticity value, SE_i is the standard error of reported income elasticity, β_1 is the coefficient that measures the existence of publication bias, Z_{ki} represent the meta-independent variables that explain the heterogeneity in reported outcomes of the literature, β_k are the coefficients that measure the effect of these variables on reported income elasticites of health care expenditure and εi is the error term.

Most of the previously cited papers which employ MRA use a general to specific (G-to-S) modeling technique, in what appears to be a general methodological consensus. This approach implies estimating a model where all variables are included, followed by the exclusion of all insignificant variables one-by-one, while checking at every step the validity of the transformations (Campos *et al.* 2005). While some scholars state that the G-to-S approach is a good tool for producing a structured model and avoiding data mining (S&D 2012, p 91), others believe that a G-to-S strategy is statistically dubious and the outcomes of such a modeling technique greatly depend on the order in which the variables have been excluded and the data has been altered (Pagan 1987).

Our multiple MRA analysis was carried out in two directions. The first one started

with the estimation of a model which included all the meta-independent variables presented in table 4.3, as in the G-to-S approach. Those insignificant moderator variables were then eliminated one at a time until only those with a p-value ≤ 0.05 remained. Correlations between variables were also taken into account when performing the eliminations. However, this does not represent our main approach, and results of the G-to-S modeling can be consulted in the Appendix D, as means of comparison with the modeling technique we preferred, which is a bottom-up approach. Starting from the simple MRA model, meta-independent variables were gradually added, checking at every phase for the significance of the newly added regressors and their effect on the joint significance of the MRA model.

Chapter 4

The data

4.1 Construction of the dataset

The final dataset is comprised of 220 elasticities gathered from 54 studies; a list of studies included in the MRA alongside the number of estimates collected from each one can be found in Appendix A. On average, a study reports 4.1 estimates, with a maximum of 20 reported estimates per study and a minimum of 1 estimate per study. The oldest research paper dates from 1977 and the most recent was published in 2013. The dataset contains studies which exhibit many differences in the methodology and data they employ. We tried to include as many of these studies as long as their outcomes could be compared or transformed into a common metric, and we coded and controlled for the observed differences in data or methods in the multiple MRA.

The database construction started from two narrative reviews of the literature on drivers of health spending: Martín *et al* (2011) and Van Elk *et al.* (2010), which were referred to previously. Regardless of the differences in scope of these two papers and the present one, the two studies proved to be very helpful for the current research, especially their references and citations, which pointed various studies that could be included in our MRA. A total of 42 studies were collected from the references of these two reviews.

In addition, a search in PubMed, Scopus, JSTOR, Google Scholar and EBSCOhost online databases using the keywords: "health expenditure", "spending", "costs" "determinants", "income", "ageing", in a variety of combinations, added 99 more studies to the data-set. Lastly, 7 additional empirical papers were discovered while reading through the references of various studies on the topic of health expenditure, so we included them as well. The results of the search are presented in Table 4.1.

As can be seen above, the initial search yielded a number of 148 total studies. Nevertheless, as previous scholars often mention as one of the shortcomings of meta-analysis, there is no guarantee that all studies on a certain topic are gathered after the search (Rose & Stanley 2005). In this particular case, the keyword search in the online databases might have failed to spot important articles which offer estimates of

Table 4.1: Data sources

Database	Number of studies
Martin Martin *et al.* paper (2011) & van Elk *et al* (2010)paper	42
PubMed	43
SCOPUS	16
EBSCOhost	19
Google Scholar	9
JSTOR	12
References	7
TOTAL	148

the effect of income on health expenditure, but do so as an auxiliary theme. Other important articles that defined their topic using a different terminology might have been overlooked as well. Thus, there is no guarantee that our initial sample of studies managed to gather all econometric investigations on the effect of income on health expenditure and therefore results should be interpreted with caution.

On the other hand, various papers identified by the search proved irrelevant for the purpose of the present study, i.e. did not employ econometric analysis, analyzed only one segment of health expenditure, such as public health expenditure or out-of-pocket payments, did not offer elasticity estimates or enough information for calculating them, did not employ income among their explanatory variables, etc. Thus, further selection criteria were implemented, in order to reduce the sample of studies to those offering comparable effect sizes, i.e elasticities. Papers which were included in the final sample of studies had to pass the following selection criteria:

(a) Use econometric methods and more specifically, regression analysis;

(b) Include income among their explanatory variables;

(c) Look at the determinants of health expenditure at a macro-level, i.e consider health care expenditure at a country/ state/ regional level, and not individual or household expenditure on health. This criteria was implemented for comparability reasons, since health expenditure decisions at individual and aggregate level are fueled by different reasons.

(d) Define as dependent variable aggregate health expenditure, and not merely a segment of it, such is public health expenditure or out-of-pocket payments. A comparison between these different types of health costs would render ambiguous results, since these entities are not comparable and different drivers might be behind their evolution.

(e) Use the double-log functional form and report clear estimates (p-values, t-values, standard errors), or otherwise, provide the appropriate statistical information needed for computing the effect size and its standard error;

(f) Are published in English .

Table 4.2: Average effect size

Average effect size	Value
Unweighted	1.0455
FEE (s.e)	1.122 (0.002)
REE (s.e)	0.921 (0.021)
95% CI (unweighted)	0.943 to 1.1578
95 % CI (FEE)	1.117 to 1.127
95 % CI (REE)	0.879 to 0.963

After applying these selection criteria, the studies which complied with all requirements and were included in our final database were in number of 54 and provided 220 elasticities. As a point of reference, Doucouliagos & Stanley (2012) found after reviewing 87 meta-analysis that, on average, a meta-analysis included 35 studies. Appendix B provides the list of all the papers not considered in our final database, alongside the reason for their exclusion.

4.2 Meta-dependent variable: the income elasticity of aggregate health care spending

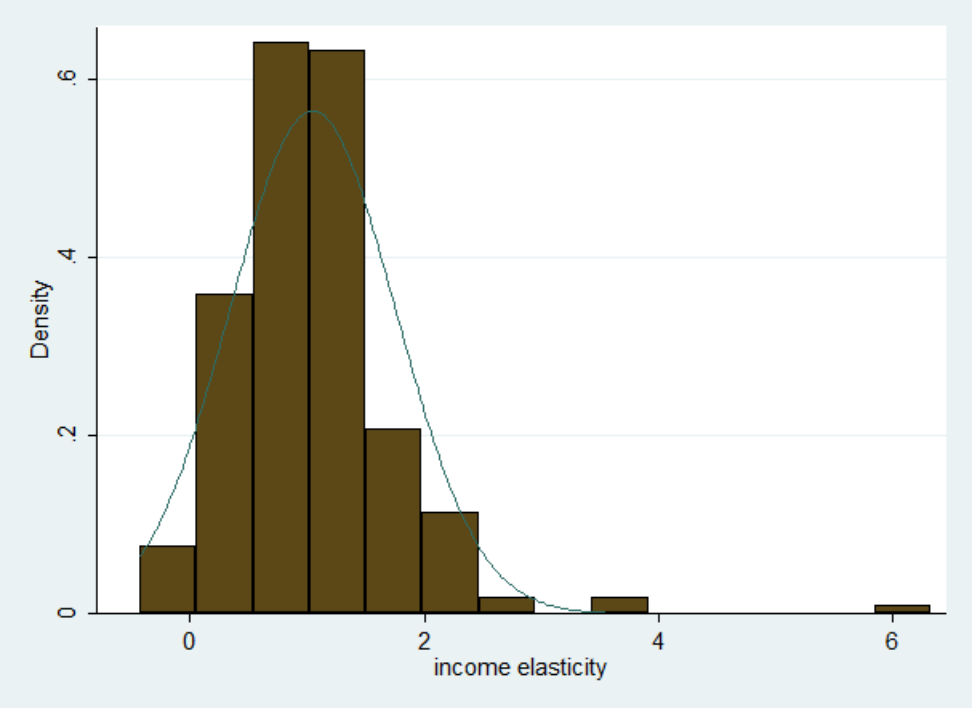
The collected income elasticities of health expenditure, as reported in the literature, vary from a minimum of -0.429 to a maximum of 6.33, with a standard deviation of 0.707. Half of the reported income elasticities lie below the value of 0.940, while the top 10 % of results are above the value of 1.68. The unweighted mean is 1.04, placing the average income elasticity close to unity. We also computed the unweighted average effect for only those elasticities corresponding to t-values situated above the 75th percentile, since higher t-values are associated with greater confidence in the estimate, and thus the resulted averaged effect size should also exhibit greater reliability. In this case, the mean value resulted to be approximately the same, at 1.050 value, with a 95 % confidence interval of (0.94, 1.15). However, as previously explained, the unweighted mean is in most cases a misleading statistic, since effect sizes have different precisions, meaning their weight in the final average effect size should vary according to this precision. Therefore, we also computed the weighted average, which provides a more robust estimate. Table 4.2 presents the values of the unweighted average, FEE and REE average effect size, therefore the average income elasticity.

It can be noticed that the weighted FEE and REE averages are different than the unweighted value. Under the FEE assumption, health care expenditure appears to be income elastic, with an elasticity of 1.122 and a low standard error of 0.002. REE average elasticity, on the other side, is slightly below unity at 0.921, which would situate health care in the necessity good category. It should be noted that it also has a higher standard error (0.021)than the FEE one, which makes it less precise.

If we would have the certainty that the elasticity estimates from which these average values have been calculated represent a random sample of all health expenditure income elasticity estimates existing in the literature, then the weighted effect would be unbiased and would represent a trustworthy measurement of the average effect size. Nonetheless, the presence of publication bias would translate into a sample of population effects which is not random, meaning that the weighted effects we have computed might be distorted, even for the REE estimate, which accounts for between-study variance. Since in fact we suspect slight publication bias in the literature, the averaged effect size cannot be considered a trustworthy representation of the outcomes reached in the literature. Instead, a more accurate representation will be the "true" effect size, which will be presented in chapter 5 among other results of the MRA analysis.

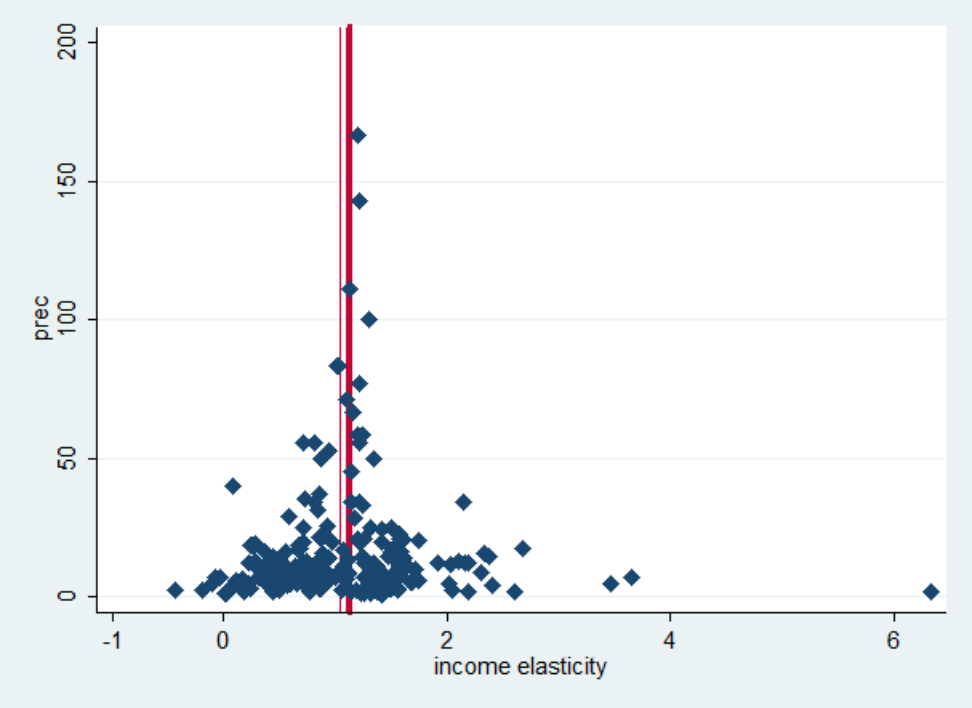
Graphic tools offer a general picture of the data and uncover possible problems, such as outliers. The frequency distribution of our effect sizes follows in certain measure a normal distribution. Considering that the estimated elasticities were extracted from regressions using the double-log function, this is somewhat expected. Most of the values lie around unity, while some elasticities exhibit values that are not in line with the majority of estimates.

Figure 4.1: Frequency distribution of income elasticities



Another visual representation of the dependent variable observations distribution was obtained by plotting the income elasticity against its precision, in a funnel plot. As it can be seen from Figure 4.2, there is a general trend of values being clustered around

Figure 4.2: Funnel graph of income elasticity of health care spending



the FEE averaged income elasticity value, which is represented by the thick line and corresponds to the value of 1.122. The thin line represents the REE averaged efect size.

Nevertheless, a number of elasticities estimates appear to be outliers and leverage points. These values were inspected, in order to discover whether they are coding errors, or actual outliers. The outliers, which signify the amounts with a low precision but very large value of the elasticity estimates, were retained in the database, since at the moment when they will be analyzed in the MRA, they will exert a low influence on the results. The reason for this is because they are weighted by measure of their precision, and since their precision is low so will be their influence.

The leverage points, on the other side, exhibit a very high precision and therefore can have a large influence on the results of the MRA. Our data seems to exhibit some values that could be considered outliers and levarage points. The papers from which these values were extracted were double checked, in search of possible coding errors. However, none were discovered, thus the values were kept, expecting for the reasons of this heterogenity to be revealed by the multiple MRA analysis.

It is obvious the funnel does not have a symmetrical shape, since values at the right of the graph seem to be missing and are more scattered than those on the left-hand side, pointing at the possibility of insignificant estimates (elasticities with low values and large standard errors) having not been reported. This visual inspection indicates the literature is affected by publication bias. Opinions on the income elasticity of

aggregate health care expenditure appear to be mixed, with roughly half of estimates (49%) reporting a value below unity, and the other half (51 %) uncovering elasticities higher than one. When looking only at the top 10 % most precise estimates, which due to their high precision are less likely to be affected by publication bias, 68 % have elasticities higher than 1, while the rest of 32 % elasticities are below unity.

4.3 Meta-independent variables

We assumed that a number of study characteristics could explain the heterogeneity in health care spending income elasticities reported by the literature:

- **degree of aggregation of data**. A number of scholars have suggested the income elasticity of health care expenditure behaves differently, depending on the level at which is analyzed. Getzen (2000) found that health care is a necessity at the individual level and behaves more like a luxury good at the national level. Furthermore, there have been some voices in the academic environment stating that performing comparisons of health care expenditure at the international level could be problematic, due to the institutional differences in health care systems, different levels of development between countries and different types of governance. Our MRA includes a dummy variable accounting for the level at which the primary study performs the analysis, whether between-countries or regions inside a country. We expect studies developed at a lower level of aggregation will report lower values of health spending income elasticity.

- **country/ region analyzed by the primary study**. Many papers looking at the aggregate health expenditure drivers have focused on OECD countries. There are however some studies in our database that use African countries, Middle Eastern or Economic Cooperation Organization countries as the base of their analysis. Our MRA will include a dummy variable of the area being studied, in order to see whether it influences in any way the reported income elasticities of health spending. We expect a positive estimated coefficient, which would suggest that analysis of OECD countries delivers higher income elasticities and reports that health care behaves as a luxury good in developed countries .

- **alternative ways in which health expenditure is measured**. The majority of studies focused on uncovering the drivers of total health costs has been measuring these costs in levels. New research suggests that attention should be given not only to the level of health expenditure, but also to the drivers of health spending growth rates. Some evidence of β-convergence in health spending has been found, where countries that initially have a low expenditure record increased growth in health spending, while those countries with an initial level of health spending already high exhibit a lower growth trend (Barros 1998). We include a dummy variable for the way in which health spending is measured, in an attempt to identify whether levels or growth rates of health expenditure influence the magnitude of the reported income elasticity.

- **type of data used**. Our MRA will control also for the data set being employed by the primary study. The purpose is to see whether employing cross-sectional, time-series or panel data yields different estimates of income elasticity. Employing panel data is expected to generate lower income elasticity estimates.

- **estimation technique employed**. Different estimation methods have been used in the literature of the determinants of HCE expenditure, from simple OLS to WLS, GLS or cointegration regression using dynamic OLS, CCEMG estimators, ECM or FMOLS estimators. The multivariate MRA will control for the method employed in the primary studies and quantify how the pick of different estimatation techniques affects the reported outcomes.

- **year of publication of study**. The purpose of controlling for the year of publication of a study is to observe whether reported values of income elasticity of HE are different with the passing of time and whether there exists a cyclical trend in the research of drivers of health care expenditure. For this purpose, we also include in the analysis the square value of the study's year of publication. If these two variables prove jointly significant and the relationship presents a concave shape, then the "economics research cycle hypothesis" can be confirmed (Havránek 2012, p.21). This hypothesis assumes that initial results in a field of research tend to be confirmed by following studies, but with the pass of time, opposite results become preferable, since they bring a new argument to the table and keep the research alive (Goldfarb 1995 in Havránek 2012).

- **moderator variables included in the primary study**. Our multivariate MRA will include a number of dummy variables that will account whether income is the sole explanatory variable used in the primary study; whether the primary research controls for institutional variables (type of health care system; physicians working as gatekeepers; no. of physicians per 1,000 people; degree of decentralization of health care system etc.) and whether the primary study controls of medical technology advancements (by including among its explanatory variables various proxies for medical technology: medical R&D, time-trend, no. of medical procedures/ 1000 people, etc). The purpose is to reveal whether employing models with different explanatory variables causes heterogeneity in reported effects of income on health care expenditure. We expect that, generally, the inclusion of other independent variables apart from income will result in lower income elasticity values.

- **conversion method of health care expenditure**. A number of studies in the literature draw attention on how different ways of converting health care expenditure in international comparisons can affect one study's outcomes. Studies have been employing PPPs, exchange rates, CPI or GDP deflator. Our MRA will control for these different conversion ways and see whether they contribute to the heterogeneity of income elasticities of health spending.

- **number of years covered by primary study data**. Research performed over a short period a time might provide unreliable results. We have decided to control for

Table 4.3: Meta-independent variables

Moderator variable	Definition	Summary statistic
SE	standard error of reported income elasticity	0.164 (0.16)
pubyear	year of publication of study(base year=1977)	24.36(7.02)
sqpubyear	square of study's year of publication	
years	number of years covered by primary data	19.60(15.59)
only_inc	=1 if income the sole explanatory variable of primary study	62
growth	=1 if health expenditure is measured in growth rates	17
panel	=1 if estimate comes from panel data	114
cross	=1 if estimate comes from cross-sectional data	80
timeseries	=1 if estimate comes from time-series data	40
OECD	=1 if estimate comes from OECD data	212
institut	=1 if model includes institutional variables	53
technology	=1 if model includes proxiesfor technology advancements	38
OLS	=1 if OLS estimation used	131
other estimation	= 1 if other estimation technique employed (WLS, GLS, ML)	42
cointeg	=1 if cointegrating regression used	64
PPP	= 1 if Purchasing Power Parity used to convert health expenditure	177
internat	=1 if international comparison of health expenditure performed	192

Note: For SE, pubyear and years, the reported summary statistic is the mean and the standard deviation. For the dummy variables, we report the number of observations for which the dummy variable is equal to 1.

this detail as well and record its influence on our meta-dependent variable.

Table 4.3 provides an overview of all meta-independent variables employed in our multivariate MRA, alongside some summary statistics.

Roughly 90 % of the studies in the meta-dataset analyze OECD countries. Approximately 56% of the primary research employs OLS techniques, while 75% of the studies performing international comparisons of health spending use PPP for converting the different income and health care expenditure values into a common metric. On average, the number of years covered by the data of a primary study is 20 years, with a minimum of one year for cross-sectional studies and a maximum of 44 years. About half of the literature (48.5%) employs panel data, while the other half uses cross-sectional or time-series data.

The correlation matrix among all independent variables can be found in Appendix C. There are few strong correlations and many of them are expected. The dummy variable accounting for the use of cross-sectional data by a primary study appears highly correlated with the number of years of the study's data ($r = -0.81$) and with the dummy accounting for a primary study that uses panel data ($r = -0.70$). Since the variables "cross', "panel" and " timeseries" are categorical variables referring to the same characteristic of a primary study, one would have to be disregarded from our multiple MRA model, and since "panel" and "timeseries" variables do not present any high correlations with other variables, the meta-independent variable "cross" will be

the one left out of the model.

The same principle was applied for the "cointeg" variable, which signals whether a primary study uses cointegration regression as its estimation technique. "OLS", "cointeg" and "other estimation" are categorical variables, and from the three of them, "cointeg" presents a high positive correlation with both "years" variable ($r = 0.60$) and "OLS" variable ($r = -0.67$), as expected. Since "OLS" and "other estimation" have lower correlations with the rest of the meta-independent variables, "cointeg" will not be included in the multiple MRA model.

The correlation relations among the other independent variables do not indicate very strong relations that could represent a threat for our multiple MRA model's reliability. It would be worth noticing the positive correlation between "PPP" and "pubyear" variables ($r=0.43$), indicating that the use of exchange rates as conversion method between international health expenditures is a rather obsolete method, and newer research tends to use PPP instead. The publication year of a study is also relatively highly correlated in a negative direction with the variable accounting for a study using international data ($r=-0.38$). This suggests that, with the pass of time, studies became more aware of the difficulties and biases the comparisons between countries with different health care systems and underlying conditions could bring, and prefer to focus on analysis of national-level data. Finally, an interesting observation is related to the correlation between the standard error of health spending's income elasticity estimate and the "cross" and "panel" variables. The magnitude of the correlation relations is roughly the same, but in opposite direction, suggesting that the use of cross sectional data generates higher standard errors ($r= 0.20$), while the use of panel data reports estimates with lower standard errors ($r=-0.20$) and therefore higher precision.

Chapter 5

Results

5.1 Testing for publication bias

The identification of causes of heterogeneity in reported health spending income elasticity values started with a check for publication bias. Due to publication selection, reported estimates can be heavily distorted, and estimating the true effect without considering the possibility that published results might have been purposely picked would still generate biased estimates of the true income elasticity.

The visual inspection of the income elasticity values plotted against their precision(Figure 4.2) revealed the presence of, what seemed, not very strong but still present, publication bias. The simple FAT-PET MRA mentioned in equation 3.7 was further estimated, to formally test for publication bias. In order to avoid the risk of heteroskedasticity, each estimate was weighted by the inverse of its squared precision. Various estimation techniques have been used, so as to ensure the robustness of our results. The outcomes of the simple MRA can be viewed in Table 5.1.

The first column corresponds to the simple WLS model. The model reports that publication bias does exist ($\hat{\beta}_1 = -1.29$), at a significance level of 5%. It also suggests, with 99% confidence, that there exists a true effect of income on health spending situated above unity value.

Due to the possible dependence between estimates collected from the same study, we continue our analysis with a cluster robust version of the WLS model, where observations are clusterd at study level. As it is expected, the coefficients' magnitudes remain the same, but the standard errors change and render the intercept insignificant, which means that we can no longer reject the null hypothesis that there is no publication selection among our collected elasticities. The intercept is still significant at 1% level.

In order to capture all aspects of our dataset, we further estimate a number of multilevel models. Since there might be a consistent within study variance, a fixed-effects model was estimated to account for this variance. The results are presented in column 3. In this case, the slope parameter has a negative sign, but continues being

Table 5.1: Simple MRA results

	(1) WLS	(2) WLS cluster robust	(3) FEML	(4) REML	(5) Median	(6) ME	(7) PEESE
$\hat{\beta}_1$(pub.bias)	-1.291**	-1.291	-0.280	0.115	-1.967***	0.0858	
	(0.529)	(1.345)	(0.417)	(0.248)	(0.640)	(0.247)	
$\hat{\beta}_1$(PEESE test)							-1.443
							(0.947)
$\hat{\beta}_0$ (true eff.)	1.151***	1.151***	1.092***	1.042***	1.133***	1.047***	1.125***
	(0.0297)	(0.0299)	(0.0688)	(0.0919)	(0.0472)	(0.0951)	(0.0290)
Observations	220	220	220	220	220	220	220

Standard errors in parantheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 = weighted least squares model, where variables have been weighted by inverse of squared precision. Column 2 = cluster robust standard errors, with clusters at study level. Column 3 = fixed effects multilevel model. Column 4 = random-effects multilevel model. Column 5 = estimation at median level. Column 6 = mixed effects multilevel model. Column 7 = independent variable is no longer the standard error, but the variance.

insignificant ($\hat{\beta}_1 = -0.28$). The significance of $\hat{\beta}_1$ at a 1.092 value suggests there exists a true effect of income elasticity, as might be expected since publication bias has not been proved in this case, and that according to this value, health care lies in the luxury good category.

Although not recommended in MRA analysis, we estimated a random-effects model as well (column 4). As might already be known, the basic assumption of a random effects-model is that the unobserved values of its dependent variable are independent of the explanatory variables. In our case, we are testing for publication bias because in fact we suspect the reported income elasticities of aggregate health expenditure are correlated with their standard error, and implicitly with the inverse of the standard error, which then violates the RE assumption. Due to this reason, S & D (2012) advocate the use of fixed-effects models in MRA, whose assumptions are respected in a higher percentage. After estimating both models, the Hausman test was implemented, in order to reveal which of these two models is preferred. With a p-value of 0.0006, the Hausman test clearly rejects the random-effects model, which proves the recommendations of the MRA literature.

The multilevel mixed effects estimation (column 6), which can successfully be used in the case of clustered observations of a single variable and unbalanced data, and is actually recommended in MRA analysis (S & D 2012, p.100), delivers a still insignificant slope parameter, strengthening the idea that the literature evaluating the income elasticity of aggregate health does not suffer majorly from publication bias.

We also estimated two models that take into consideration the variables' measure of central tendency. The results of the model using the median values of each of the two variables included in the simple MRA are reported in column 5 and they offer a different view than previous estimations. Since the median is a measure of central tendency, it

is not as seriously affected by outliers as the mean. In this model, the publication bias indicator becomes highly significant, meaning that the null hypothesis of no publication bias can be rejected at 1 % confidence level. Its negative sign would suggest that preference for publication is given to those studies reporting income elasticity values of health expenditure below unity. In what regards the true income elasticity $\hat{\beta}_1$, it exhibits a value close to unity, at 1.04.

A final attempt to tackle the problem of the likely dependence among our effect sizes was performed by running an MRA on the weighted averages of each study's estimate and their standard errors (results not reported in Table 5.1). In this case, each study provides a single averaged effect size of all its estimates. Therefore, the number of observations of the MRA is equal to the number of studies (54, compared to 220 as in the previous models). Each estimate of income elasticity has been weighted by the sample size. Afterwards, the average effect size for each study was computed. The same strategy was adopted for calculating average standard errors and t-values. The estimation generated an insignificant $\hat{\beta}_1 = -0.070$ and a $\hat{\beta}_0 = 1.12$ significant at 10%. We cannot reject the null of no publication bias once again, while the true effect value is significant and for the first time lies slightly below unity ($\hat{\beta}_1$=0.966). While employing the average estimate of each study is considered to offer a more realistic evaluation of the MRA statistical significance (S &D 2012, p.72), it should not be overlooked that the much smaller number of observations employed in this case gives the model less degrees of freedom and makes it unable to capture the entire bias resulted from the collection of different estimates from a same study, even though they have been weighted by the sample size. This could explain, if even only a part, the newly found below-unity true income elasticity value.

We have presented various models used to test for the existence of publication selection in the literature on the influence of income per capita on aggregate health care expenditure. The various results seem to indicate, in general terms, that publication bias is not a serious problem. Many of the employed specifications did not find significant values for the slope parameter. Those that have, found negative values of $\hat{\beta}_1$, which would indicate that reported estimates of income elasticity of aggregate health care expenditure are skewed towards values lower than unity. This suggests a tendency in the literature towards publishing those studies which find that health care behaves as a necessity rather than a luxury good. If we consider the suggestion of S &D (2012, p.159) that larger values of $\hat{\beta}_1$ are indicators of larger publication selection, then the idea of under-reporting elasticities higher than one is mildly supported by the magnitudes of $\hat{\beta}_1$, which are no bigger than the value of -1.96.

Most evidence drives the conclusion towards the idea that the hypothesis of no publication bias cannot be rejected. The FE-cluster robust model and ME model, by employing all observations, providing robust standard errors and taking into consider-

ation the unbalanced nature of data, could also be considered reliable, and they both point at estimates not significantly biased by publication selection.

In what regards the existence of a true estimate of income elasticity, it could be said that a general consensus has been uncovered. All the models found significant $\hat{\beta}_0$ estimates, which in all cases were above unity. This finding points to a true effect of income on aggregate health care spending, suggesting at the same time that health expenditure increases at a slightly faster rate than income does, and thus, health care can be labeled a luxury good. The result comes logically, since when there is no significant publication bias, the results reported by the literature can actually be trusted and averaging them should provide a trustworthy measure of the true income elasticity.

Since the PET test rejected the null hypothesis, which assumed that there is no true effect of income on health expenditure, we also estimated β_0 using the PEESE test, which is believed to offer a better estimation of the corrected true effect. It should be remembered that the PEESE model – equation 3.9 – considers as independent variable in the regression not the standard error of the estimates, but their variance, and that the model has no intercept. Results of the PEESE test are presented in column 7. The true effect estimate stands at 1.12 value, thus supporting the idea of health care being a luxury good.

All in all, it could be said that publication bias does not represent a serious concern for the literature analyzing the effect of per capita income on health spending. The weak evidence that was found of its presence suggests a mild preference for under-reporting elasticities above unity. This would point to some efforts made towards convincing governments to take more responsibility in ensuring that all citizens benefit from health services, since with an income elasticity below one, health care becomes a necessity. In fact, as results of our PEESE test suggest, health care is more a luxury good and, as Newhouse (1977, p.123) emphasized, its main purpose might be directed rather towards "caring than curing". The finding also supports Kotzian's (2003, p.24) discovery that individuals tend to evaluate and be satisfied with health care systems performance based on their 'caring' ability, while the 'curing' function is often taken for granted.

Although no strong evidence for publication bias could be discovered, there is still heterogeneity among the reported health care spending income elasticities which needs to be explained. We further look into more detail at the possible factors causing this heterogeneity by employing the multivariate MRA model.

5.2 Explaining heterogeneity. Multiple MRA

The results of the G-to-S modeling rendered significant 6 independent variables, that passed the robustness check and maintained their significance across different specifications. However, since the general-to-specific approach is rather frowned upon in the econometric world, we present the results of our parsimonious strategy only in Appendix

D. As the main approach towards the multiple MRA, we adopt an opposite direction, starting from the bivariate FAT-PET model and gradually enriching the model with more independent variables.

Although some newly added variables prove insignificant, the joint significance testing we perform at each step does not show at any time that the model becomes jointly insignificant. Thus, our final specification includes all 14 independent variables. It can be observed that the variable "cross", which accounts for cross-sectional data having been used in the primary studies, alongside the variable expressing if cointegration was the estimation technique used in the primary studies, were not included in the multiple MRA, due to high correlations with other independent variables.

The simple WLS MRA (column 1 of Table 5.2)is able to explain 48 % of the heterogeneity in reported income elasticity values. The Breusch-Pagan test does not indicate any signs of heteroskedasticity among the variables ($\text{chi2}_{14} = 394.34$, p-value < 0.0001). Ramsey's RESET test rejects the null hypothesis that the model has no omitted variables ($F_{3,199} = 3.03, \text{p-value} = 0.03$), which brings questions about other variables not included in our model, which could be able to explain the heterogeneity in collected elasticity values. We try to adress this issue by employing the mixed effects multilevel model in column 5.

The first estimation of the multiple MRA model reveals a number of 7 significant variables. At a 1% significance level, the results show that studies which analyze OECD countries report income elasticities of aggregate health spending higher with 0.50% than analysis focusing on non-OECD nations.

Furthermore, accounting for institutional factors in one's research will decrease the values of the income elasticity obtained by 0.45 %. Including proxies for medical technology advancements in the primary model will also give lower income elasticities than research which ignores these factors, the decrease being of 0.20%.

Using PPPs instead of exchange rates for conversion of health expenditures of different countries will generate income elasticities higher with 0.31%. Interestingly, the literature seems to exhibit a time trend as well. Studies having been published more recently, appear to report lower income elasticities of aggregate health expenditure. More exactly, newer research will report income elasticity values 0.01% lower than results provided by papers published a year before. Although the magnitude of the change brought by time is not very big, it is still an interesting fact to notice that time seems to have an influence on the results found in the literature of health care spending determinants.

Significant at 5% level, the variable accounting for the use of time series by a primary study reveals that reported elasticities will be 0.26% higher than those of studies employing cross-sectional data. The result brings about questions, since cross-sectional data is mainly used by initial studies in the area, and they are the ones which normally report the highest health costs income elasticity values. The reliability of the time series variable is proven doubtful by following specification, as although it remains significant,

Table 5.2: Multiple MRA results

	(1) WLS	(2) WLS	(3) WLS cluster robust	(4) ME	(5) PEESE
S.E	-0.355	0.436	-0.355	0.0930	
	(0.658)	(0.647)	(1.110)	(0.248)	
OECD	0.507***	0.498***	0.507***	0.368*	0.508***
	(0.103)	(0.0978)	(0.179)	(0.221)	(0.103)
PPP	0.319***	0.194***	0.319***	0.141	0.319***
	(0.0630)	(0.0653)	(0.116)	(0.159)	(0.0630)
institut	-0.455***	-0.447***	-0.455***	-0.238**	-0.466***
	(0.0934)	(0.0888)	(0.156)	(0.106)	(0.0920)
internat	0.0744*	0.0628	0.0744	0.170	0.0719*
	(0.0409)	(0.0390)	(0.0806)	(0.194)	(0.0408)
growth	-0.135	-0.185	-0.135	-0.0247	-0.154
	(0.149)	(0.142)	(0.112)	(0.142)	(0.146)
years	-0.000288	-0.00131	-0.000288	-0.00663	-0.000268
	(0.00247)	(0.00236)	(0.00475)	(0.00566)	(0.00248)
panel	-0.127	-0.122	-0.127	-0.0862	-0.0983
	(0.105)	(0.100)	(0.132)	(0.173)	(0.0959)
timeseries	0.266**	0.309**	0.266	0.691***	0.289**
	(0.127)	(0.121)	(0.215)	(0.225)	(0.122)
OLS	0.0429	0.0428	0.0429	-0.116	0.0500
	(0.0546)	(0.0519)	(0.0627)	(0.125)	(0.0534)
other_est	0.110*	0.125**	0.110	-0.143	0.114*
	(0.0651)	(0.0620)	(0.0848)	(0.150)	(0.0650)
only_inc	0.0461	0.0445	0.0461	-0.0147	0.0469
	(0.0318)	(0.0302)	(0.0287)	(0.0964)	(0.0318)
technology	-0.206***	-0.211***	-0.206**	-0.208 **	-0.210***
	(0.0675)	(0.0642)	(0.101)	(0.106)	(0.0673)
pubyear	-0.0165***	0.0828***	0.0828***	-0.0140	-0.0165***
	(0.00404)	(0.0212)	(0.0397)	(0.0111)	(0.00404)
sqpubyear		-0.00213***	-0.00213***		
		(0.000447)	(0.0008)		
se_sq					0.239
					(1.784)
Constant	0.760***	-0.207	0.760**	1.032**	0.721***
	(0.180)	(0.265)	(0.304)	(0.439)	(0.169)
R^2	0.48	0.53	0.48		0.48
Observations	220	220	220	220	220

Dependent variable: income elasticity of aggregate health care spending. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

it experiences major shifts in the coefficient's magnitude (for example, in the ME model, the coefficient's magnitude is 0.691, while in the WLS specification is 0.266). The initial set of results also uncovers that research on the effect of per capita income on health care costs which is done at an international level, using data from different countries, generates higher outcomes by 0.07% than those studies employing national-level data. However, the significance of this variables is low, with a p-value of 0.071.

Column 2 tests the hypothesis of the "economics research cycle". For this, the square of the primary study's year of publication is added to the model. This newly added variable, although is has a very low magnitude (coefficient = -0.002), is significant at 1% level. In addition, the variable accounting for the study's publication year remains significant and changes sign, exhibiting a coefficient of 0.08. Thus, we cannot reject the hypothesis that the literature on the determinants of health care expenditure follows an "economics research cycle". This signifies that there seem to be trends in reported research to favor a groundbreaking finding for a period of time, but after some years give preference to outcomes which contradict the initial findings and become the new trend in research.

Although heteroskedasticity was not signaled as a problem for the WLS model, we further estimated the model employing cluster robust standard errors, with estimates clustered at the study level (column 3), in order to check for other problems that could bias our results, such as autocorrelation. The magnitude of the coefficients is unchanged, as normal, but some independent variables lose their significance. More specifically, the variables accounting for international data and time series being used in the primary studies lose their significance. Furthermore, the publication year of a study seems not to exert any influence on the reported health costs income elasticities. The intercept keeps its significance, suggesting the existence of a true elasticity value, while the coefficient of the standard error remains insignificant, as in the previous specifications, offering no evidence for publication bias.

As a further robustness check, the regression model is estimated by ways of multilevel mixed-effects, following the recommendations of various scholars (Bateman & Jones 2003, Stanley & Doucouliagos 2012), in order to account for the dependence among estimates and correct the standard errors adequately. Only 3 independent variables, alongside the intercept, are significant in this case. The variable which codes studies that analyze OECD countries keeps exerting a positive influence on reported income elasticities, exhibiting a coefficient of 0.36. The true effect remains significant, with an increased value than in previously estimated models ($\hat{\beta}_0 = 1.03$), which would bring health care closer to the luxury good category, although its income elasticity so close to unity does not offer very strong support. Papers which include institutional variables and proxies accounting for medical technological developments tend to generate lower income elasticities of health care spending, while the publication year of a study is still insignificant.

Lastly, column 6 presents the results of the PEESE model, employed to get the better estimates model of our true effect. When using the square standard error as an independent variable, instead of the simple standard error as was the case in the previous columns, the true income elasticity is estimated at 0.721, with a 99 % confidence. The slope of the squared standard error, which is the sign for publication bias, is insignificant (p- value = $-.89$), meaning that publication selection is not present in the literature in great measure.

5.2.1 The "best practice" estimate

Having uncovered some of the causes underlying the heterogeneity in reported income elasticities of health care spending, we can further estimate the true income elasticity, dependent on the study's components. Whether a study uses OLS or cointegration techniques, cross sectional or panel data or observations spanning over 5 or 20 years are details that influence research outcomes, and we can asses in what measure. Furthermore, since we already know some of the research's characteristics that have an effect on its outcomes, we can consider as"ideal study" the ones which take into consideration all variables found significant by the multiple MRA.

In an initial phase, no preference or conditions were set over what an "ideal study" should look like. Thus, the independent variables were all considered at their sample means. The estimated income elasticity for this case resulted to be significant at a 1 % level, with a value of 0.92 and a 95 % confidence interval $(0.73, 1.10)$. Compared to the average effect size computed in Table 4.2, this result is very similar to the REE average.

Next, we estimate the true underlying income elasticity by placing conditions on the variables found significant across the multiple specifications in Table 5.2. We thus set the dummy variables "OECD", "institut" and "tech" at their maximum sample value, which is 1. This means that we compute the estimate generated by those studies that analyze OECD countries and account for institutional and medical technical advancements. The rest of the variables are left at their sample mean value. The resulted "best practice" estimate is significant at 1 % level, with a value of 0.51 and a 95 % confidence interval of $(0.16, 0.85)$. This would mean that studies believed to include all factors that might influence reported income elasticites of health spending and thus being the"best practice" studies, generate outcomes which point at health care being a necessity, and not a luxury good. This finding is rather striking if we compare it to the average effect size, which is more vague about the true nature of health care, having values close to unity.

If the assumptions change and we also put a condition on the"time series" and "PPP" variables, which were found significant across most of our multiple MRA specifications, then the estimate of the true income elasticity rises to 0.67, significant at 1% level with a 95% confidence interval of (0.26, 1.08). Once again, health care can be classified as a necessity, with an 10% increase in income levels (measured as GDP/capita in most

studies) bringing about an increase of 7% in aggregate health spending.

However, the "best practice" definition is rather a subjective one and most certainly other variables or structural factors of research should be included among the assumptions, as proof standing the R^2 of the multiple MRA, which is able to explain only about 48% of the heterogeneity in reported elasticities.

Chapter 6

Discussion and conclusion

The present study has tried to uncover the true magnitude of the relation between per capita income and aggregate health care expenditure. Previous research in the field has not been able to reach a general consensus and during time there have been voices stating that health care is a necessity good, while others advocated for rather labeling it as a luxury one.

Our research reveals that the truth lies somewhere in the middle and that the MRA analysis of the literature is sensitive to the estimation method used. We found very little evidence of publication bias, and the one that was discovered indicates that values above unity of the income's effect on health expenditure tend to be slightly under-reported or under-published, and studies which find that health care is a necessity are given preference for publication. In other words, the distribution of income elasticities is skewed to the left. This idea is also supported by the visual representation of our observations, which resembles a slightly asymmetrical funnel plot. The conclusion makes sense in the context of simple MRA, as reported"true" income elasticity of health spending stands at values above unity, which is higher than the REE averaged value. The difference between the two is a result of negative publication bias, which instead of reporting all income elasticity estimates, picks smaller ones, causing the averaged elasticity reported by the literature to be smaller than the true one. Once more variables are added to the MRA model, the coefficient indicating publication bias becomes insignificant across all specifications, and the null hypothesis of no publication bias cannot be rejected. However, one must not neglect the fact that not being able to reject the null hypothesis does not automatically mean there is absolutely no publication bias in the literature. We thus conclude that while publication bias is not a serious concern for the literature, we suspect it exerts an influence on outcomes in a negative direction. Further research, including more observations, possibly from the studies we had to discard from our dataset due to various reasons (see Appendix B), could bring stronger evidence.

Our research finds strong support for the existence of certain study characteristics which have an influence on the magnitude of reported health spending income elasticity.

More specifically, papers which include among their independent variables a proxy that accounts for institutional characteristics of the health care systems under analyze, e.g. number of physicians per 1,000 persons, degree of centralization, social insurance coverage etc., will generate lower income elasticity estimates. The same applies for those studies that account for the technological developments in the medical field, by introducing in their regression model either a time-trend, a measure of medical R & D or other appropriate proxies. The type of data used by a study also exerts an influence on its resulting income elasticity, with studies employing time series data reporting higher income elasticity values. Moreover, we find that research focused on analyzing OECD countries generates higher elasticity values.

These findings have a significant importance for future research, as they point at a number of components that should be included in a study, in order to get closer to the"best practice" research design. Thus, future studies on the determinants of health care expenditure should account for institutional and technological advancements in the medical field and include a measure of these among the independent variables. They should also consider employing time-series or panel data, instead of cross-sectional one. Moreover, the use of PPP instead of exchange rates when carrying out the conversion among health expenditure levels expressed in various national currencies also seems to be worthy of consideration, since the "PPP" variable's coefficient managed to pass all robustness checks, being insignificant only in the mixed-effects multilevel model.

Although we expected the type of estimation technique employed by a research (OLS or cointegration) to have a significant influence on reported income elasticities, this assumption has not been proven by the MRA outcomes. Instead, the hypothesis of an "economic research cycle" was confirmed, meaning that the literature on the determinants of health care expenditure follows certain cyclical trends when reporting its results. Therefore, while we did not find that the literature suffers from considerable publication bias, there still exists another influence that affects reported outcomes, and this bias is not due to publishers, but rather to authors, who try to comply with the trends of research of the time. This would imply that the lower income elasticities which started being reported in the last decade are not only a result of methodological advancements, but also a phase in the research cycle, corresponding to refuting previous results and adopting a skeptical view. The trends in research which were identified in section 1.1 offer support to the confirmed "economic research cycle hypothesis".

In what regards the true income elasticity size, the multiple MRA finds across different specification that a true effect of income on aggregate health spending does exist, but its value is fluctuating: a WLS cluster robust model suggests that the income elasticity is below unity level (0.896), while the ME estimate is slightly above one (1.064), being more in line with the results provided by the simple MRA. We would conclude that if we consider the weak evidence for publication bias, then the REE average elasticity size (0.921) could, in general lines, be a reliable representation of the true income elasticity.

Slightly higher values of the true effect could be considered as well, corresponding to the situation of publication bias working towards the publishing of under-inflated income elasticities. Then, the results provided by the ME multilevel model (1.032), which also accounts for unobserved between-studies heterogeneity and the unbalanced nature of our data, would represent a reliable expression of the true income's elasticity value.

This results, however, while statistically significant, do not seem to provide an answer to the initial question. With an income elasticity close to unity, health care cannot be named neither a luxury good, nor a necessity. Its fluctuating true underlying values would seem to offer support to Getzen's similar claims, who mentions that health care cannot be put into either category of goods, but it is both a luxury and a necessity, as income elasticity varies with the level of analysis (Getzen 2000). He reveals that at an individual level, health care stands in the necessary good category, while at higher aggregated level, i.e. national, it behaves as a luxury good. This is an interesting hypothesis to test by future MRA research. Estimates observed at individual level could be included among the meta-observations, although it is open to discussion whether outcomes measured at individual level could be compared with those computed at a national level, since the decisions regarding the level of health care spending are based on widely different factors at the two levels.

Another interesting topic for further research would be to check whether and how different political forces influence reported estimates of income elasticity. Our hypothesis could be tested by upcoming research by including an independent variable in the multiple MRA model which accounts for left-wing or right-wing governance and determines whether it has any influence on health spending's income elasticity. This suggestion arises from the political considerations attached to the categorization of health care as a necessity or luxury. It would be expected that a left- wing governance, which promotes equal access to health for all citizens, would push towards results that present health care as a necessity. On the other side, right-wing parties in power would prefer empirical evidence of health care being a luxury good, which would then allow market forces to intervene and pass more of the health provision and funding to the private sector.

Bibliography

Allen, M. (2009): "Meta-Analysis." *Communication Monographs* 76 (4): pp. 398-407.

Ang, J. (2010): "The Determinants of Health Care Expenditure in Australia." *Applied Economics Letters* 17 (7).

Atella, V. & G. Marini (2006): "Is Health Care Expenditure Really a Luxury Good? Re-Assessment and New Evidence Based on OECD Data." *SSRN Scholarly Paper*. Rochester, NY: Social Science Research Network.

Baltagi, Badi H. & F. Moscone (2010): "Health Care Expenditure and Income in the OECD Reconsidered: Evidence from Panel Data." *Economic Modelling* 27 (4): pp. 804-11.

Barros, P. (1998): "The Black Box of Health Care Expenditure Growth Determinants." *Health Economics*7(6): pp. 533-44.

Beaman, A. L. (1991): "An Empirical Comparison of Meta-Analytic and Traditional Reviews" *Personality and Social Psychology Bulletin* 17 (3): pp. 252-57.

Bech, M., T. Christiansen, E. Khoman, J. Lauridsen, & M. Weale (2011): "Ageing and Health Care Expenditure in EU-15." *The European Journal of Health Economics* 12 (5): pp. 469-78.

Bel, G. & X. Fageda (2009): "Factors Explaining Local Privatization: A Meta-Regression Analysis." *Public Choice* 139 (1-2): pp. 105-19.

Bilgel, F. & Kien C. T. (2013): "The Determinants of Canadian Provincial Health Expenditures: Evidence from a Dynamic Panel" *Applied Economics* 45 (2): pp. 201-12.

Blomqvist, A G. & R. A. L. Carter (1997): "Is Health Care Really a Luxury" *Journal of Health Economics* 16 (2): pp. 207-29.

Borenstein, M., L. V. Hedges, J. PT Higgins, & H. R. Rothstein (2011): "Introduction to Meta-Analysis." John Wiley & Sons.

Campos, J., N. R. Ericsson, & D. F. Hendry (2005): "General-to-specific Modeling: An Overview and Selected Bibliography". Board of Governors of the Federal Reserve System. *International Finance Discussion Papers* Number 838.

Chernichovsky, D. & S. Markowitz (2004): "Aging and Aggregate Costs of Medical Care: Conceptual and Policy Issues." *Health Economics* 13 (6): pp. 543-62.

Christiansen, T., M. Bech, J. Lauridsen, & P. Nielsen (2006): "Demographic Changes and Aggregate Health-Care Expenditure in Europe." *ENEPRI Research report.*

Clemente, J., C. Marcuello, A. Montañés, & F. Pueyo (2004): "On the International Stability of Health Care Expenditure Functions: Are Government and Private Functions Similar?" *Journal of Health Economics* 23 (3): pp. 589-613.

Culyer, A. J. (1989): "The Normative Economics of Health Care Finance and Provision." *Oxford Review of Economic Policy* 5 (1): pp. 34 58.

De Linde L. M., T. D. Stanley, & H. Doucouliagos (2013): "Does the UK Minimum Wage Reduce Employment? A Meta-Regression Analysis." *British Journal of Industrial Relations*, August

Di Matteo, L. (2003): "The Income Elasticity of Health Care Spending: A Comparison of Parametric and Nonparametric Approaches." *The European Journal of Health Economics* 4 (1): pp. 20-29.

Dormont, B., M. Grignon, & Hélène Huber (2006): "Health Expenditure Growth: Reassessing the Threat of Ageing." *Health Economics* 15 (9).

Doucouliagos, H. & T. D. Stanley (2009): "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* 47 (2): pp. 406-28.

Dreger, C., & H. E. Reimers (2005): "Health care expenditures in OECD countries: A panel unit root and cointegration analysis." *IZA Discussion paper series* 1469.

Eliot, T. S (1952): "Complete Poems and Plays 1909-1950." New York: Harcourt

Brace.

Feng J. (2011): "Macro Determinants of Health Expenditure in China." *Fudan Journal of the Humanities & Social Sciences* 4 (3): pp. 18-37.

Freeman, D. G. (2003): "Is Health Care a Necessity or a Luxury? Pooled Estimates of Income Elasticity from US State-Level Data." *Applied Economics* 35 (5).

Fries, J. F. (1983): "Aging, natural death, and the compression of morbidity." *The Milbank Memorial Fund Quarterly.* Health and Society 61 (3): 397.

Gbesemete, K. P., & U. G. Gerdtham (1992): "Determinants of Health Care Expenditure in Africa: A Cross-Sectional Study." World Development 20 (2): pp. 303-8.

Gerdtham, U. G., B. Jönsson, M. MacFarlan, & H. Oxley (1998): "Factors Affecting Health Spending: A Cross-Country Econometric Analysis. " in OECD (1995b) "New Directions in Health Care Policies: Improving Cost Control and Effectiveness." *Developments in Health Economics and Public Policy* 6: pp. 113-34.

Gerdtham, U.G. (1992): "Pooling International Health Care Expenditure Data." *Health Economics* 1 (4): pp. 217-31.

Gerdtham, U. G. & B. Jönsson (1991): "Price and Quantity in International Comparisons of Health Care Expenditure." *Applied Economics* 23 (9): pp. 1519-28.

Gerdtham, U. G. & B. Jönsson (1992): "International Comparisons of Health Care Expenditure – Conversion Factor Instability, Heteroscedasticity, Outliers and Robust Estimators." *Journal of Health Economics* 11 (2): pp. 189-97.

Gerdtham, U. G. & M. Löthgren (2000): "On Stationarity and Cointegration of International Health Expenditure and GDP." *Journal of Health Economics* 19 (4): pp. 461-75.

Gerdtham, U. G., J. Sogaard, F. Andersson, & B. Jönsson (1992): "An Econometric Analysis of Health Care Expenditure: A Cross-Section Study of the OECD Countries." *Journal of Health Economics* 11 (1): 63-84.

Getzen, T. E. (1992): "Population Aging and the Growth of Health Expenditures." *Journal of Gerontology* 47 (3): pp. S98-104.

Glass, Gene V. 1976. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher*, pp. 3- 8.

Greenberg, D. H., C. Michalopoulos, & P. K. Robins (2003): "Meta-Analysis of Government-Sponsored Training Programs, A." *Indus. & Lab. Rel. Rev.* 57: 31.

Hansen, P. & A. King (1996): "The Determinants of Health Care Expenditure: A Cointegration Approach." *Journal of Health Economics* 15 (1): pp. 127-37.

Hartwig, J. (2008): "What Drives Health Care Expenditure?–Baumol's Model of 'Unbalanced Growth' Revisited." *Journal of Health Economics* 27 (3): pp. 603-23.

Hartwig, J. (2011): "Can Baumol's Model of Unbalanced Growth Contribute to Explaining the Secular Rise in Health Care Expenditure? An Alternative Test." *Applied Economics* 43 (2).

Havránek, T. (2014): "Meta-Analysis in International Economics." Accessed May 13. available online at: [http://samba.fsv.cuni.cz/ phd/Thesis_D_Havranek.pdf.]

Havránek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* 85 (2): pp. 234-44.

Herwartz, H. & B. Theilen (2003): "The Determinants of Health Care Expenditure: Testing Pooling Restrictions in Small Samples." *Health Economics* 12 (2).

Hitiris, T & J Posnett (1992): "The Determinants and Effects of Health Expenditure in Developed Countries." *Journal of Health Economics* 11 (2): pp. 173-81.

Hitiris, T. (1997): "Health Care Expenditure and Integration in the Countries of the European Union." *Applied Economics* 29 (1).

Karatzas, G. (2000): "On the Determination of the US Aggregate Health Care Expenditure." *Applied Economics* 32 (9).

Kiymaz, H., Y. Akbulut, & A. Demir (2006): "Tests of Stationarity and Cointegration of Health Care Expenditure and Gross Domestic Product: An Application to Turkey." *The European Journal of Health Economics* 7 (4): pp. 285-89.

Lago-Peñas, S., D. Cantarero-Prieto, & C. Blázquez-Fernández (2013): "On the Re-

lationship between GDP and Health Care Expenditure: A New Look." *Economic Modelling* 32 (May): pp. 124- 29.

Le Pen, Y. (2002): "An International Comparison of Health Care Expenditure Determinants." *International Conferences on Panel Data.*

Leu, R. E. (1986): "The Public-Private Mix and International Health Care Costs." *Public and Private Health Services*, pp. 41-63.

Longhi, S., P. Nijkamp, & J. Poot (2006): "The Impact of Immigration on the Employment of Natives in Regional Labour Markets: A Meta-Analysis." *IZA Discussion Papers.* Available online at:http://www.econstor.eu/handle/10419/33654. Last accessed [17 March 2014].

López-Casasnovas, G.& M. Saez (2007): "A Multilevel Analysis on the Determinants of Regional Health Care Expenditure: A Note." *The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care* 8 (1): pp. 59-65.

Martin, J. J. M., M. P. Lopez del Amo Gonzalez, & M. D. Cano Garcia (2011): "Review of the Literature on the Determinants of Healthcare Expenditure." *Applied Economics* 43 (1): pp. 19-46.

McCoskey, S. K. & T. M. Selden (1998): "Health Care Expenditures and GDP: Panel Data Unit Root Test Results." *Journal of Health Economics* 17 (3): pp. 369-76.

Mehrara, A. A. F. & A. R. Fazaeli (2009): "The Relationship between Health Expenditures and Economic Growth in Middle East & North Africa (MENA) Countries." *Faslname-Modiraite Salamat* 35 (12): pp. 49-59.

Milne, R. & H. Molana (1991): "On the Effect of Income and Relative Price on Demand for Health Care: EC Evidence." *Applied Economics* 23 (7): pp. 1221-26.

Mosca, I. (2007): "Decentralization as a Determinant of Health Care Expenditure: Empirical Analysis for OECD Countries." *Applied Economics Letters* 14 (7).

Murillo, C., C. Piatecki,& M. Saez (1993): "Health Care Expenditure and Income in Europe." *Health Economics* 2 (2).

Murthy, N. R. & A. Okunade (2000): "Managed Care, Deficit Financing, and Aggregate Health Care Expenditure in the United States: A Cointegration Analysis." *Health Care*

Management Science 3 (4).

Murthy, V. N. R. & V. Ukpolo (1994): "Aggregate Health Care Expenditure in the United States: Evidence from Cointegration Tests." *Applied Economics* 26 (8): 797.

Murthy, V. N. R. & A. A. Okunade (2009): "The Core Determinants of Health Expenditure in the African Context: Some Econometric Evidence for Policy." *Health Policy* 91 (1): pp. 57-62.

Murthy, V. N. R. (1992): "Conversion Factor Instability in International Comparisons of Health Care Expenditure: Some Econometric Comments." *Journal of Health Economics* 11 (2): pp. 183-87.

Newhouse, J. P. (1977): "Medical-Care Expenditure: A Cross-National Survey." *The Journal of Human Resources* 12 (1): pp. 115-25.

O' Connell, J. M. (1996): "The Relationship between Health Expenditures and the Age Structure of the Population in OECD Countries." *Health Economics* 5 (6).

O'Neill, C., L. Groom, A. J. Avery, D. Boot, & K. Thornhill (2000): "Age and Proximity to Death as Predictors of GP Care Costs: Results from a Study of Nursing Home Patients." *Health Economics* 9 (8): pp. 733-38.

Okunade, A. A. (2005): "Analysis and Implications of the Determinants of Healthcare Expenditure in African Countries." *Health Care Management Science* 8 (4): pp. 267-76.

Okunade, A. A. & M. C. Karakus (2001): "Unit Root and Cointegration Tests: Time-series versus Panel Estimates for International Health Expenditure Models." *Applied Economics* 33 (9).

Okunade, A.A & V. N.R. Murthy (2002): "Technology as a 'major driver' of health care costs: A cointegration analysis of the Newhouse conjecture." *Journal of Health Economics* 21 (1): pp. 147-59.

Okunade, A., M. Karakus, & C. Okeke (2004): "Determinants of Health Expenditure Growth of the OECD Countries: Jackknife Resampling Plan Estimates." *Health Care Management Science* 7 (3).

Pagan, A. (1987): "Three econometric methodologies: a critical appraisal." *Journal of Economic Surveys* 1 (1-2): pp. 3-23.

Pammolli, F., M. Riccaboni, & L. Magazzini (2012): "The Sustainability of European Health Care Systems: Beyond Income and Aging." *The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care* 13 (5): pp. 623-34.
Parkin, D., A. McGuire, & B. Yule (1987): "Aggregate Health Care Expenditures and National Income: Is Health Care a Luxury Good?" *Journal of Health Economics* 6 (2): pp. 109-27.

Phillips, J. M. & E. P. Goss (1995): "The Effect of State and Local Taxes on Economic Development: A Meta-Analysis." *Southern Economic Journal* 62 (2): 320.

Phillips, P. C. B. (1986): "Understanding Spurious Regressions in Econometrics." *Cowles Foundation Discussion Paper* 757. Cowles Foundation for Research in Economics, Yale University.

Roberts, J. (1999): "Sensitivity of Elasticity Estimates for OECD Health Care Spending: Analysis of a Dynamic Heterogeneous Data Field." *Health Economics* 8 (5).

Roberts, J. (2000): "Spurious Regression Problems in the Determinants of Health Care Expenditure: A Comment on Hitiris (1997)." *Applied Economics Letters* 7 (5).

Rose, A. K., & T. D. Stanley (2005): "A Meta-Analysis of the Effect of Common Currencies on International Trade." *Journal of Economic Surveys* 19 (3): pp. 347-65.

Rudner, L. M., G.V. Glass, D. L. Evartt (2002): "Meta-Stat: A User's Guide to the Meta-Analysis of Meta-Stat." ERIC Clearinghouse on Assessment and Evaluation Department of Measurement, Statistics and Evaluation, University of Maryland. Available online at: http://www.edres.org/meta. Last accessed [07 April 2014].

Samadi, A. & E. H. Rad (2013): "Determinants of Healthcare Expenditure in Economic Cooperation Organization (ECO) Countries: Evidence from Panel Cointegration Tests."

Sen, A. (2005): "Is Health Care a Luxury? New Evidence from OECD Data." *International Journal of Health Care Finance and Economics* 5 (2): pp. 147-64.

Seshamani, M. & A. Gray (2004): "Ageing and Healthcare Expenditure: The Red Herring Argument Revisited." *Health Economics* 13 (4).

Stanley, T. D. (2001): "Wheat from Chaff: Meta-Analysis as Quantitative Literature

Review." *Journal of Economic Perspectives*,pp 131-50.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* 70 (1): pp. 103-27.

Stanley, T. D., C. Doucouliagos, & S. B. Jarrell (2008): "Meta-Regression Analysis as the Socio-Economics of Economics Research." *The Journal of Socio-Economics* 37 (1): pp. 276-92.

Stanley, T. D. & H. Doucouliagos (2012): "Meta-Regression Analysis in Economics and Business." Vol. 5. Routledge.

Stanley, T. D. & S. B. Jarrell (1989): "Meta-Regression Analysis: A Quantitative Method of Literature Surveys." *Journal of Economic Surveys* 3 (2): pp. 161-70.

Stanley, T. D. & S. B. Jarrell (1998): "Gender Wage Discrimination Bias? A Meta-Regression Analysis." *Journal of Human Resources*, pp. 947-73.

Stearns, S. C. & E. C. Norton (2004): "Time to Include Time to Death? The Future of Health Care Expenditure Predictions." *Health Economics* 13 (4).

Sülkü, S. N. & A. Caner (2011): "Health Care Expenditures and Gross Domestic Product: The Turkish Case." *The European Journal of Health Economics* 12 (1): pp. 29-38.

Thornton, J. A. & J. L. Rice (2008): "Determinants of Healthcare Spending: A State Level Analysis." *Applied Economics* 40 (22).

Uman, L. S. (2011): "Systematic Reviews and Meta-Analyses." *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 20 (1): pp. 57-59.

Van Elk, R., E. Mot, & P. H. Franses (2010): "Modeling Healthcare Expenditures: Overview of the Literature and Evidence from a Panel Time-Series Model." *Expert Review of Pharmacoeconomics & Outcomes Research* 10 (1): pp. 25-35.

Wang, Z. (2009): "The Determinants of Health Expenditures: Evidence from US State-Level Data." *Applied Economics* 41 (4).

Weichselbaumer, D. & R. Winter-Ebmer (2005): "A Meta-Analysis of the International Gender Wage Gap." *Journal of Economic Surveys* 19 (3): pp. 479-511.

Weil, T. P. (1994): "Comparisons of Medical Technology in Canadian, German, and US Hospitals." *Hospital & Health Services Administration* 40 (4): pp. 524-33.

Werblow, A., S. Felder, & P. Zweifel (2007): "Population Ageing and Health Care Expenditure: A School of 'red Herrings'?" *Health Economics* 16 (10).

Wilson, R. M. (1999): "Medical care expenditures and GDP growth in OECD nations." *American Association of Behavioral and Social Sciences Journal* 2, pp: 159-171.

Yavuz, N. C., V. Yilanci, & Z. A. Ozturk (2013): "Is Health Care a Luxury or a Necessity or Both? Evidence from Turkey." *The European Journal of Health Economics* 14 (1): pp. 5-10.

Zweifel, P., S. Felder, & M. Meiers (1999): "Ageing of Population and Health Care Expenditure: A Red Herring?" *Health Economics* 8 (6): pp. 485-96.

Appendices

APPENDIX A

Studies included in final dataset

Author(s)	Title	Year pub.	Obs.
Ang J.	"The determinants of health care expenditure in Australia"	2010	2
Atella& Marini	"Is Health Care Expenditure Really a Luxury Good? Re-Assessment and New Evidence Based on OECD Data"	2006	1
Baltagi& Moscone	"Health care expenditure and income in the OECD reconsidered: Evidence from panel data"	2010	6
Barros	"The black box of health care expenditure growth determinants"	1998	5
Bilgel& Tran	"The determinants of Canadian provincial health expenditures: evidence from a dynamic panel"	2013	2
Blomqvist& Carter	"Is health care really a luxury?"	1997	6
Chernichovsky& Markowitz	"Aging and aggregate costs of medical care: conceptual and policy issues"	2004	1
Christiansen *et al.*	"Demographic changes and aggregate health-care expenditure in Europe"	2006	4
Clemente *et al.*	"On the international stability of health care expenditure functions: are government and private functions similar?"	2004	17
Di Matteo	"The Income Elasticity of Health Care Spending: A Comparison of Parametric and Nonparametric Approaches"	2003	2
Dreger& Reimers	"Health Care Expenditures in OECD Countries : A Panel Unit Root and Cointegration Analysis"	2005	3
Feng	"Macro Determinants of Health Expenditure in China"	2011	3
Freeman	"Is health care a necessity or a luxury? Pooled estimates of income elasticity from US state-level data"	2003	5
Gbesemete& Gerdtham	"Determinants of health care expenditure in Africa: A cross-sectional study"	1992	3
Gerdtham *et al.*	"Factors affecting health spending: a cross-country econometric analysis", in "New Directions in Health Care Policies: Improving Cost Control and Effectiveness"	1998	2
Gerdtham	"Pooling international health care expenditure data"	1992	5
Gerdtham& Jönsson	"Price and quantity in international comparisons of health care expenditure"	1991	2

Continued on next page

Appendix A: Studies included in the dataset

Author(s)	**Title**	**Year pub.**	**No. estimates**
Gerdtham& Jönsson	"International comparisons of health care expenditure–conversion factor instability, heteroscedasticity, outliers and robust estimators"	1992	12
Gerdtham *et al.*	"An econometric analysis of health care expenditure: A cross-section study of the OECD countries"	1992	7
Getzen	"Population aging and the growth of health expenditures"	1992	7
Herwartz& Theilen	"The determinants of health care expenditure: testing pooling restrictions in small samples"	2003	20
Hitiris& Posnett	"The determinants and effects of health expenditure in developed countries"	1992	2
Hitiris	"Health care expenditure and integration in the countries of the European Union"	1997	4
Karatzas	"On the determination of the US aggregate health care expenditure"	2000	1
Kiymaz *et al.*	"Tests of Stationarity and Cointegration of Health Care Expenditure and Gross Domestic Product: An Application to Turkey"	2006	2
Lago-Peñas *et al.*	"On the relationship between GDP and health care expenditure: A new look"	2013	5
Le Pen	"An International Comparison of Health Care Expenditure Determinants"	2002	8
Leu	"The public-private mix and international health care costs"	1986	1
López-Casasnovas& Saez	"A multilevel analysis on the determinants of regional health care expenditure: a note"	2007	1
Mehrara *et al.*	"The Relationship between Health Expenditures and Economic Growth in Middle East & North Africa (MENA) Countries"	2012	1
Milne& Molana	"On the effect of income and relative price on demand for health care: EC evidence"	1991	3
Mosca	"Decentralization as a determinant of health care expenditure: empirical analysis for OECD countries"	2007	1
Murillo *et al.*	"Health care expenditure and income in Europe"	1993	4
Murthy& Okunade	"Managed care, deficit financing, and aggregate health care expenditure in the United States: A cointegration analysis"	2000	1

Continued on next page

Appendix A: Studies included in the dataset

Author(s)	**Title**	**Year pub.**	**No. estimates**
Murthy& Ukpolo	"Aggregate health care expenditure in the United States: Evidence from cointegration tests"	1994	1
Murthy& Okunade	"The core determinants of health expenditure in the African context: some econometric evidence for policy"	2009	4
Murthy	"Conversion factor instability in international comparisons of health care expenditure: Some econometric comments"	1992	4
Newhouse	"Medical-Care Expenditure: A Cross-National Survey"	1977	2
O' Connell	"The relationship between health expenditures and the age structure of the population in OECD countries"	1996	5
Okunade *et al.*	"Determinants of Health Expenditure Growth of the OECD Countries: Jackknife Resampling Plan Estimates"	2004	2
Okunade	"Analysis and implications of the determinants of healthcare expenditure in African countries"	2005	10
Okunade& Karakus	"Unit root and cointegration tests: timeseries versus panel estimates for international health expenditure models"	2001	1
Okunade& Murthy	"Technology as a 'major driver' of health care costs: a cointegration analysis of the Newhouse conjecture"	2002	1
Pammoli *et al.*	"The sustainability of European health care systems: beyond income and aging"	2012	6
Parkin *et al.*	"Aggregate health care expenditures and national income: Is health care a luxury good?"	1987	2
Roberts	"Sensitivity of elasticity estimates for OECD health care spending: analysis of a dynamic heterogeneous data field"	1999	4
Roberts	"Spurious regression problems in the determinants of health care expenditure: a comment on Hitiris (1997)"	2000	3
Samadi& Homaie Rad	"Determinants of Healthcare Expenditure in Economic Cooperation Organization (ECO) Countries: Evidence from Panel Cointegration Tests"	2013	1
Sen	"Is health care a luxury? New evidence from OECD data"	2005	9
Sulku& Caner	"Health care expenditures and gross domestic product: the Turkish case"	2011	1
Thornton& Rice	"Determinants of healthcare spending: a state level analysis"	2008	6

Continued on next page

Appendix A: Studies included in the dataset

Author(s)	**Title**	**Year pub.**	**No. estimates**
Van Elk *et al.*	"Modeling healthcare expenditures: overview of the literature and evidence from a panel time-series model"	2010	1
Wang	"The determinants of health expenditures: evidence from US state-level data"	2009	12
Yavuz *et al.*	"Is health care a luxury or a necessity or both? Evidence from Turkey"	2013	1

APPENDIX B

Studies not included in final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Abbas F. & U. Hiemenz (2013)	"What determines public health expenditures in Pakistan? Role of income, urbanization and unemployment"	dep. var.: public HE
Abolhallaje M. *et al.* (2013)	"Determinants of catastrophic health expenditure in Iran"	measures contribution of HE to impoverishment (in the regression model, HE is an explanatory variable)
Alcalde-Unzu J. *et al.* (2009)	"Cross-country disparities in health-care expenditure: a factor decomposition"	does not employ regression analysis
Ariste R. & J. Carr (2003)	"New Considerations on Empirical Analysis of the Determinants of Canadian Provincial Government Health Expenditures, 1966-1998"	dep. variable: governmental HE
Ayé R. *et al.* (2011)	"Factors determining household expenditure for tuberculosis and coping strategies in Tajikistan"	measures contribution of HE to impoverishment
Bech M. *et al.* (2011)	"Ageing and health care expenditure in EU-15"	offers no t-values or standard error values for the computed long-run elasticities
Benjamin A.E (1986)	"Determinants of state variations in home health utilization and expenditures under Medicare"	dep. var.: Medicaid expenditure, not aggregate US HE
Blanco-Moreno A. *et al.* (2013)	Public healthcare expenditure in Spain: measuring the impact of driving factors	dep. var.: public HE
Breyer F. & S. Felder (2006)	"Life expectancy and health care expenditures: A new calculation for Germany using the costs of dying"	prediction; no estimates of HE income elasticity
Brinda E. M. *et al.* (2012)	"Nature and determinants of out-of-pocket health expenditure among older people in a rural Indian community"	dep. var.: out-of-pocket HE
Buczko W. (1986)	"Physician utilization and expenditures in a Medicaid population"	does not refer to aggregate HE

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Cantarero D. (2005)	"Decentralization and health care expenditure: the Spanish case"	dep. var.: public HE
Cantarero D. & S. Lago-Peñas (2010)	"The determinants of health care expenditure: a re-examination"	dep. var.: public HE
Cantarero D. & S. Lago-Peñas (2012)	"Decomposing the determinants of health care expenditure: the case of Spain"	dep. var.: public HE
Carrion-i-Silvestre J. (2005)	"Health care expenditure and GDP: Are they broken stationary?"	focuses on testing for stationarity; does not offer estimates of HE income elasticity;
Chang S. *et al.* (2013)	"The determinants of health care expenditure towards the end of life: evidence from Taiwan"	considers health costs ocurred only in last years of life
Chernew M. (2010)	"Health care spending growth: can we avoid fiscal Armageddon?"	does not employ econometric analysis
Cid Pedraza C. & L. Prieto Toledo (2012)	"Out-of-pocket health spending: the case of Chile, 1997 and 2007"	dep. var.: out-of-pocket HE
Crivelli L. *et al.* (2006)	"Federalism and regional health care expenditures: an empirical analysis for the Swiss cantons"	dep. var.: public + insurance HE
De Meijer C. *et al.* (2011)	"Determinants of long-term care spending: age, time to death or disability?"	income not include in model
De Meijer C. *et al.* (2013)	"Health expenditure growth: looking beyond the average through decomposition of the full distribution"	dep. var.: acute & pharmaceutical HE; no estimation for aggregate HE
Di Matteo L. (2010)	"The determinants of the public-private mix in Canadian health care expenditures: 1975-1996"	dep. var.: public HE, private HE; no estimation for aggregate HE
Di Matteo L. (2005)	"The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time"	dep. var.: provincial HE, personal HE

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Di Matteo L. (2009)	"Policy choice or economic fundamentals: what drives the public-private health expenditure balance in Canada?"	dep. var.: public share of Canadian total HE
Di Matteo L. ()	"Physician numbers as a driver of provincial government health spending in Canadian health policy"	dep.var.: public HE
Di Matteo L. (1998)	"Evidence on the determinants of Canadian provincial government health expenditures: 1965-1991"	dep. var.: public HE
Dicker M. & J.H. Sunshine (1998)	"Determinants of financially burdensome family health expenses: United States, 1980"	dep.var.: household HE
Dormont B. *et al.* (2006)	"Health expenditure growth: reassessing the threat of ageing"	focuses on relationship between HE and ageing; income not included in the model
Filmer D. & L.Pritchett (1999)	"The impact of public spending on health: does money matter?"	dep.var.: health status
Fujiwara Y. *et al.* (2000)	"Regulatory factors of medical care expenditures for older people in Japan-analysis based on secondary medical care areas in Hokkaido"	only looks at health costs for older people
Furuoka F. (2011)	"What Are the Determinants of Health Care Expenditure? Empirical Results from Asian Countries"	dep. var.: public HE
Gerdtham U. G. & M. Löthgren (2002)	"New panel results on cointegration of international health expenditure and GDP"	only tests for cointegration; offers no estimates values
Gerdtham U.G. & B. Jönsson (2000)	" International comparisons of health expenditure: Theory, data and econometric analysis"	reviews previous literature; no new information
Gerdtham U. G. & M. Löthgren (2000)	"On stationarity and cointegration of international health expenditure and GDP"	only tests for cointegration; offers no estimates values
Getzen T. ()	"Health care is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures"	no regression analysis

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Gianoni M. & T. Hitiris (2002)	"The regional impact of health care expenditure: the case of Italy"	dep. var.: public HE
González-González C. *et al.* (2011)	"Health care utilization in the elderly Mexican population: expenditures and determinants"	dep. var.: health care utilization
Gordon R. L. *et al.* (1997)	"Determinants of US local health department expenditures, 1992 through 1993"	dep. var.: public HE
Han K.*et al.* (2013)	"Determinants of health care expenditures and the contribution of associated factors: 16 cities and provinces in Korea, 2003-2010"	dep. var.: health insurance HE
Hansen P. & A. King (1996)	"The determinants of health care expenditure: A cointegration approach"	only test for stationarity between HE and GDP per capita; no estimates of HE income elasticity offered
Hartwig J. (2008)	"What drives health care expenditure?–Baumol's model of 'unbalanced growth' revisited"	dep.var.: total current HE = total HE - total investment in medical facilities
Hartwig J. (2011)	"Can Baumol's model of unbalanced growth contribute to explaining the secular rise in health care expenditure? An alternative test"	dep.var.: total current HE = total HE - total investment in medical facilities
Hung J. H. & L. Chang (2008)	"Has cost containment after the National Health Insurance system been successful? Determinants of Taiwan hospital costs"	dep.var.: hospital costs
Jacobs K. (1989)	"Determinants of income and expenditure development in the annual reports of the health insurance of the Federal Republic of Germany"	dep. var.: public HE
Jerret M. *et al.* (2003)	"Environmental Influences on Healthcare Expenditures: An Exploratory Analysis from Ontario, Canada"	income not among the independent variables
Jewell T. *et al.* (2003)	"Stationarity of health expenditures and GDP: evidence from panel unit root tests with heterogeneous structural breaks"	new tests for stationarity, but no estimates of HE determinants

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Jowett M. (1999)	"Bucking the trend? Health care expenditures in low-income countries 1990-1995"	does not employ econometric analysis
Kornai J. & J. McHale (2000)	"Is Post-Communist Health Spending Unusual?"	dep. var.: share of public HE in total HE
Kronenberg C. & P.P. Barros ()	"Catastrophic healthcare expenditure - Drivers and protection: The Portuguese case"	dep. var.: public HE
Lakshmi T. S. *et al.* (2012)	"An Analysis of Pattern and Determinants of Public Expenditure on Health in Andhra Pradesh, India"	dep.var.: public HE
Li. Y *et al.* (2012)	"Factors affecting catastrophic health expenditure and impoverishment from medical expenses in China: policy implications of universal health insurance"	focused on measuring the contribution of health expenditure to impoverishment
Liu Y. *et al* (2003)	"Medical expenditure and rural impoverishment in China"	research measures poverty impact of health expenditure
Lu X. & M. Wu (2010)	"Determinants of household healthcare expenditure of rural floating population in Beijing: a Tobit model approach"	dep.var.: household HE
Magazzino C. & M. Mele (2012)	"The Determinants of Health Expenditure in Italian Regions"	dep. var.: household HE
McCoskey S. & T. M. Selden (1998)	"Health care expenditures and GDP: panel data unit root test results"	only offer results for unit root results; no estimate of HE income elasticity
Mehrara M. & G. Sharzei (2012)	"A Study of the Relationship between Health Expenditure and Environmental Quality in Developing Countries"	language: Arabic
Morgan S. & C. Cunningham (2011)	"Population aging and the determinants of healthcare expenditures: the case of hospital, medical and pharmaceutical care in British Columbia, 1996 to 2006"	does not employ regression analysis
Moscone F. & E. Tosetti (2010)	"Health expenditure and income in the United States"	dep. var.: personal HE

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Ashar M.M & A. S. Shah Iqbal (2012)	"Socio-economic determinants of household out-of-pocket payments on healthcare in Pakistan"	dep.var.: household out-of-pocket HE
Murthy N. (2004)	"Health care expenditures in Africa: An econometric analysis"	uses same data and offers the same results as Murthy & Okunade (2009)
Nabyonga Orem J. *et al.* (2013)	"Health care seeking patterns and determinants of out-of-pocket expenditure for malaria for the children under-five in Uganda"	dep. var.: out-of-pocket HE
Okunade A. & C. Surradetcha (2000)	"Health care expenditure inertia in the OECD countries: A heterogeneous analysis"	dep.var.: HE inertia = one-period lagged medical care expenditure
Okunade A. *et al.* (2010)	"Determinants of Thailand household healthcare expenditure: the relevance of permanent resources and other correlates"	dep. var.: household HE
Omotor D. (2009)	"Determinants of Federal Government Health Expenditures in Nigeria"	dep. var.: public HE
Onwujekwe O. *et al.* (2010)	"Investigating determinants of out-of-pocket spending and strategies for coping with payments for healthcare in southeast Nigeria"	dep. var.: out-of-pocket HE
Pan J. & G. Liu (2012)	"The determinants of Chinese provincial government health expenditures: evidence from 2002-2006 data"	dep. var.: public HE
Parker S.W. & R. Wong (1997)	"Household income and health care expenditures in Mexico"	dep. var: household HE
Potrafke N. (2010)	"The growth of public health expenditures in OECD countries: Do government ideology and electoral motives matter?"	dep. var.: public HE
Rahman T. (2008)	"Determinants of public health expenditure: some evidence from Indian states"	dep. var.: public HE
Reeves A. *et al.* ()	"The political economy of austerity and healthcare: Cross-national analysis of expenditure changes in 27 European nations 1995-2011"	dep. var.: public HE

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Reich O. *et al.* (2012)	"Exploring the disparities of regional health care expenditures in Switzerland: some empirical evidence"	dep. var. is made of components of HE that represent only 56 % of total HE
Reinhardt U. E. (2003)	"Does The Aging Of The Population Really Drive The Demand For Health Care?"	does not employ econometric analysis
Rivera B. L. Currais (1999) &	"Income Variation and Health Expenditure: Evidence for OECD Countries"	in the model, health investment is an independent variable & the paper analyzes its influence on human development (dep.var.)
Rous J. J. & D. R. Hotchkiss (2003)	"Estimation of the determinants of household health care expenditures in Nepal with controls for endogenous illness and provider choice"	dep. var.: household HE
Seshamani M. & A. Gray (2004)	"Ageing and healthcare expenditure: the red herring argument revisited"	focuses on relation between HE and ageing; income not included in model
Sesma-Vázquez S. *et al.* (2005)	"Catastrophic health expenditures in Mexico: magnitude, distribution and determinants"	language: Spanish
Solakoglu E. G. & A. Civan (2012)	"Does morbidity matter? Perceived health status in explaining the share of healthcare expenditures"	dep.var.: share of GDP in HE
Stearns S. C. & E.C. Norton (2004)	"Time to include time to death? The future of health care expenditure predictions"	focuses on the HE levels of different age categories older than 65 years; income not included in regression model
Su T. T. *et al.* (2006)	"Catastrophic household expenditure for health care in a low-income society: a study from Nouna District, Burkina Faso"	HE enters the model as explanatory variable to impoverishment levels
Su T.T *et al.* (2006)	"Determinants of household health expenditure on western institutional health care"	dep. var.: household HE

Continued on next page

Appendix B: Studies excluded from final dataset

Author(s)& Year pub.	**Title**	**Reason for exclusion**
Subramanian S. V. *et al.* (2002)	"The macroeconomic determinants of health"	dep.var.: health status, not HE
Tokita T. (2000)	" Healthcare expenditure and the major determinants in Japan"	decomposes in (non) elderly/inpatient-outpatient expenditure, no aggregate measure
Toor I. A. & M. S. Butt (2005)	"Determinants of health care expenditure in Pakistan "	dep. var.: public HE
Vatter A. & C. Rüefli (2003)	"(measures public/ private HE seperately) Do Political Factors Matter for Health Care Expenditure? A Comparative Study of Swiss Cantons"	dep. var.: public HE & private HE seperately
Wade M. & S. Berg (1995)	"Causes of Medicaid expenditure growth"	dep.var.: Medicaid HE, not aggregate levels
Weisbrod B. A. (1991)	"The Health Care Quadrilemma: An Essay on Technological Change, Insurance, Quality of Care, and Cost Containment"	does not employ econometric analysis
Werblow A. *et al.* (2007)	"Population ageing and health care expenditure: a school of 'red herrings'?"	focuses on relation between HE and ageing; income not included in model
You X. & Y. Kobayashi (2011)	"Determinants of out-of-pocket health expenditure in China: analysis using China Health and Nutrition Survey data"	dep. var.: out-of-pocket HE
Yu T. H. K. & H. Y. Chu (2007)	"Is health care really a luxury? A demand and supply approach"	does not employ log-log functional form
Zweifel P. *et al.* (1999)	"Ageing of population and health care expenditure: a red herring?"	focuses on relationship between HE and different categories of population older than 65 years; income not included in model

Note: HE=health expenditure; dep.var.= dependent variable of study's regression model

APPENDIX C

Correlation matrix

Variables	S.E	OECD	PPP	institut	internat	growth rates	years	panel	timeseries	cointeg	other estimate	onlyinc	tech	pubyear	cross
S.E	1.00														
OECD	0.07	1.00													
PPP	-0.06	-0.08	1.00												
institut	0.08	-0.05	0.08	1.00											
internat	0.08	-0.03	-0.11	-0.39	1.00										
growth rates	0.14	0.09	0.07	-0.07	0.09	1.00									
years	-0.11	0.33	0.21	-0.29	0.11	0.33	1.00								
panel	-0.20	0.17	0.15	-0.07	0.16	0.23	0.45	1.00							
time series	0.01	0.11	0.10	-0.10	-0.14	-0.03	0.41	-0.43	1.00						
cointeg	0.06	0.15	0.15	-0.27	0.03	0.15	0.60	0.08	0.45	1.00					
other est	-0.12	0.05	0.11	0.28	-0.20	-0.09	-0.12	0.17	-0.18	-0.27	1.00				
onlyinc	0.02	0.11	-0.24	-0.32	0.21	-0.14	0.03	-0.23	0.33	0.11	-0.12	1.00			
tech	0.05	0.14	0.10	-0.15	0.07	0.29	0.36	0.40	-0.16	0.41	-0.08	-0.27	1.00		
pubyear	-0.16	-0.15	0.43	0.23	-0.38	-0.01	0.29	0.30	0.06	0.25	0.13	-0.26	0.17	1.00	
cross	0.20	-0.28	-0.24	0.16	-0.06	-0.21	-0.81	-0.69	-0.32	-0.44	-0.03	-0.01	-0.29	-0.36	1.00
OLS	0.04	-0.17	-0.22	0.03	0.14	-0.07	-0.44	-0.19	-0.27	-0.67	-0.49	-0.01	-0.29	-0.32	0.41

APPENDIX D

Multiple MRA results

General to specific modeling

	(1) General WLS	(2) Preferred WLS	(3) WLS	(4) Cluster robust	(5) ME	(6) PEESE
se_inc	-0.355	-0.151	0.543	-0.151	0.129	
	(0.658)	(0.564)	(0.574)	(0.983)	(0.241)	
OECD	0.507***	0.493***	0.479***	0.493**	0.239	0.495***
	(0.103)	(0.103)	(0.1000)	(0.190)	(0.214)	(0.180)
PPP	0.319***	0.322***	0.217***	0.322***	0.124	0.322***
	(0.0630)	(0.0642)	(0.0677)	(0.103)	(0.156)	(0.103)
institut	-0.455***	-0.410***	-0.389***	-0.410***	-0.231**	-0.420***
	(0.0934)	(0.0905)	(0.0878)	(0.146)	(0.100)	(0.126)
internat	0.0744*	0.0659*	0.0610	0.0659	0.0978	0.0642
	(0.0409)	(0.0395)	(0.0382)	(0.0629)	(0.195)	(0.0658)
growth_rates	-0.135					
	(0.149)					
years	-0.000288					
	(0.00247)					
panel	-0.127					
	(0.105)					
timeseries	0.266**	0.366***	0.392***	0.366*	0.665***	0.363*
	(0.127)	(0.0623)	(0.0606)	(0.189)	(0.181)	(0.191)
OLS	0.0429					
	(0.0546)					
other_est	0.110*					
	(0.0651)					
only_inc	0.0461					
	(0.0318)					
tech	-0.206***	-0.212***	-0.217***	-0.212*	-0.239 *	-0.216**
	(0.0675)	(0.0633)	(0.0612)	(0.108)	(0.117)	(0.106)
pubyear	-0.0165***	-0.0184***	0.0668***	-0.0184*	-0.0202**	-0.0184*
	(0.00404)	(0.00393)	(0.0222)	(0.0103)	(0.00990)	(0.0103)
pubyear_sq			-0.00183***			
			(0.000470)			
se_sq						0.395
						(1.709)
Constant	0.760***	0.746***	-0.103	0.746**	1.090***	0.742**
	(0.180)	(0.141)	(0.257)	(0.300)	(0.410)	(0.285)
Observations	220	220	220	220	220	220
R^2	0.48	0.44	0.45	0.44	-	0.44

Dependent variable: income elasticity of aggregate health care spending. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 1 = general MRA model, all independent variables inclued; weighted by inverse of squared precision. Column 2 = MRA model without previously resulted insignificant variables. Column 3 = square of publication year variable introduced, to test for "economics research cycle hypothesis". Column 4 = cluster robust s.e.; cluster at study level. Column 5 = mixed-effects multilevel model. Column 6 = the s.e of effect size replaced from the independent variables by its squared value.