ΠΑΝΤΕΙΟΝ ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΟΙΝΩΝΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ & ΠΕΡΙΦΕΡΕΙΑΚΗΣ ΑΝΑΠΤΥΞΗΣ

PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES
DEPARTMENT OF ECONOMIC & REGIONAL DEVELOPMENT
Πρόγραμμα Μεταπτυχιακών Σπουδών
Κατεύθυνση
Εφηρμοσμένων Οικονομικών και Διοίκησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θέμα:

Meta – analysis: Method and Applications

Κεφαλά Αγάθη

Τριμελής Επιτροπή:
Clive Richardson
Σταύρος Ντεγιαννάκης
Χρυσόστομος Στοφόρος

Αθήνα, 2017

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Μεταπτυχιακής Διπλωματικής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής. Δηλώνω, συνεπώς, ότι αυτή η Διπλωματική Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν επίσημες θέσεις.

Acknowledgements

First, I express my sincere gratitude to Professor Clive Richardson for his guidance, patience and encouragement throughout the development of the project. I sincerely thank my co-partner Victor for his understanding during the time of my post-graduate education. My thanks go also to my parents, Dimitri and Maria, for their confidence and their love during all these years. Last but not least, i would especially like to express my gratitude to my husband, Alexis, who has always supported me and helped me overcoming the difficulties, and is the one to whom I dedicate this dissertation.

Περίληψη

Ο σκοπός αυτής της διπλωματικής είναι η ανάλυση και η μελέτη της μεθόδου της μετά-ανάλυσης. Ιδιαίτερη σημασία δίνεται στο τι έχει κάνει δημοφιλή αυτή την μέθοδο καθώς και τα πλεονεκτήματα και τα μειονεκτήματα της. Ως μελέτη περίπτωσης διερευνάται η αιτιώδης σχέση μεταξύ της κατανάλωσης ενέργειας και του Ακαθάριστου Εγχώριου Προϊόντος (ΑΕΠ). Στο παρελθόν, επιστημονικά άρθρα και διπλωματικές εργασίες ασχολήθηκαν με τη σημαντικότητα αυτής της σχέσης και υπήρξαν πολλές ενδείξεις που σχετίζουν τις δυο αυτές μεταβλητές. Σε αυτή τη διπλωματική διενεργείτε μία μετά-ανάλυση με στόχο να παρατηρηθεί αν υπάρχει τελικά κάποια συσχέτιση και αν υπάρχει τι είδους κατεύθυνση παρατηρείται ότι έχει. Σε αυτή τη μετά-ανάλυση περιέχονται 158 μελέτες που καλύπτουν την χρονική περίοδο από το 1978 έως το 2011. Από την εφαρμογή των μοντέλων παλινδρόμησης μπορούμε να συμπεράνουμε ότι τα αποτελέσματα υποστηρίζουν την ένδειξη συσχέτισης μεταξύ κατανάλωσης ενέργειας και ΑΕΠ αλλά αποτυγχάνουν να δείξουν την κατεύθυνση αυτής της σχέσεως.

Λέξεις κλειδιά: κατανάλωση ενέργειας, ΑΕΠ, μετά-ανάλυση, λογαριθμική παλινδρόμηση

Abstract

The purpose of this dissertation is to analyze and study the statistical method of meta-analysis. Particular emphasis is placed on what makes this method popular and also on its strengths and weaknesses. As a case study of the method, the causal relationship between energy consumption and economic growth is investigated. The importance of identifying the direction of causality between them has been analyzed in many papers and previous dissertations. There is a variety of evidence indicating some form of relation between GDP and energy consumption. In this dissertation a meta-analysis is conducted in order to observe if there is a causal relation between them and, if there is, in which direction. 158 studies are included in this meta-analysis covering a period from 1978 to 2011. Following the application of regression models, the results support the causal relation between GDP and energy consumption but do not show the direction of this causal relation.

keywords: energy consumption, GDP, meta-analysis, economic growth, logistic regression

Περιεχόμενα

PART 1: THEORETICAL APPROACH TO META-ANALYSIS

1.1 Introduction

Meta-analysis is a statistical technique that is understood as a quantitative summary of empirical findings from studies in a specific field of interest in order to reach a general conclusion (Petitti, 2000). The observational unit is a study and the aim of meta-analysis is to identify a common statistical measure that is shared among studies and is called effect size.

Meta-analysts, through the investigation of the primary studies, examines if the effect size of a parameter is statistically significant. By combining many single studies, meta-analysis succeeds in increasing the sample size and it can be used for the investigation of similar studies under the same issues. The method of meta-analysis is based on the argument that one study alone cannot produce generalizable results. Researchers who have developed meta-analysis in order to conduct deeper analysis and combine similar studies, expect as a result a more generalized effect of the statistically significant parameters.

Supporters of this method claim that there is no other alternative and objective solution which can combine previous research findings.

Research production in medicine and in other fields of social science is rapidly growing. It is estimated that 40,000 scientific journals are being published and researchers present their studies with a frequency of one every 30 seconds, 24 hours a day, 7 days a week. The effort to reach a logical conclusion leads researchers to methods that pool the results of multiple studies. Nowadays meta-analysis is a powerful tool but it is a controversial tool as well.

1.2 History of Meta analysis

Meta-analysis began to play an important role on the scientific scene more than 50 years ago. Nevertheless, methods related to meta-analysis already existed and have been observed from the beginning of the last century. The statistician Karl Pearson is believed to be the first researcher; he collected correlation indicators from a range of studies to investigate the overall effectiveness of vaccination against measles (Pearson, 1904)

The procedure of synthesizing multiple studies has existed since the existence of research data. After Pearson, it took 40 years to Fisher (1944) to realize that generalized significant levels of independent significance tests, which individually are not all statistically significant, are higher than what one would expect because of random factors. Along with Fisher a large number of researchers such as Tippet and Cochrane, tried to combine separate studies in order to find a general result.

The term meta-analysis, introduced firstly by Gene Glass of the University of Arizona in 1976, referred to a philosophy at first, not so much to a statistical technique. He believed that a literature review should be as systematic as primary research. He said that "Meta-analysis refers to the analysis of analysis" (Glass, 1976), confirming meta-analysis as an analytical method.

Since Gene Glass, meta-analysis has become a widely accepted research tool and the first articles were published in the early 1980s (Glass, Smith 1976 - Glass, McGaw, Smith 1981 - Hunter, Schmidt, Jackson 1982). At first meta-analysis was used in medical treatment, public health and medicine and generally in the field of health. At the end of the 1980's the first books on the meta-analytic method were published and therefore the use of meta-analysis has increased. It is noteworthy that the references in this scientific field have multiplied geometrically over time. Simultaneously with the increase in studies in meta-analysis has been an increase in articles concerning the methods of meta-analysis itself. (Dickersin & Berlin, 1992)

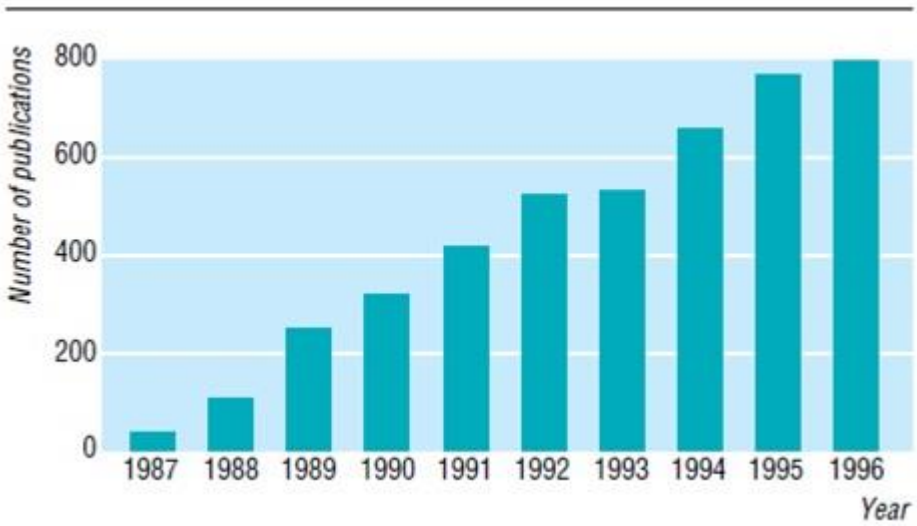

Figure 1: Number of publications about meta-analysis, from 1987 to 1996
Source: (Egger & Smith, 1997)

In 1992 the Cochrane Collaboration was founded in Oxford, taking its name from Archie Cochrane, a British researcher who contributed to the development of epidemiology as a science. He talked about methods of averaging means across independent studies and built the foundation that modern meta-analysis uses. The Cochrane Collaboration consists of a wide network of clinical and other health professionals. The foundation has spread widely through the years and now counts 37,000 contributors from more than 130 countries in Europe, North and Latin America, Africa and Australia. The purpose of this foundation is to conduct systematic reviews and meta-analysis especially in the medical field.

In social science the competent organization is the Campbell Collaboration. It promotes positive social change through the production and use of systematic reviews.

1.3 Goal and purposes of meta-analysis

Meta-analysis is a method that collects, analyzes, examines and summarizes multiple data from different but comparable studies or trials dealing with the same topic, trying to reach a general result. When the effect size (or the treatment effect) is stable between studies, meta-analysis will produce this effect. But when the effect size is different from one study to another meta-analysis will try to find and analyze the cause of differentiation (Dickersin & Berlin, 1992).

Likewise meta-analysis focuses on how much difference something makes (the magnitude of an effect) and the direction which is represented by the effect size, and is capable of overcoming small sample sizes. It examines the average effect, the precision and the consistency of effects and the relationship between study features and effects. It is an alternative method to the more traditional review methods.

Other goals of meta-analysis are to describe the distribution, its mean and establish a confidence interval around this mean; test that the mean is different from zero; explore if the studies are homogenous; and examine the relationship between study features and effect size. And finally researchers try to detect bias and heterogeneity in studies.

Meta-analysis is used by researchers in many fields of science such as education, psychology, finance, marketing and social science. In recent decades, data collection from various fields of science is growing rapidly. Researchers need help to find answers especially when a high level of uncertainty is observed when articles and reports disagree. Meta-analysis is also useful because the validity of a hypothesis cannot be asserted from the results of a single study.

1.4 Types of review

Blettner, Sauerbrei, Schlehofer, Scheuchenpflug & Friedenreich (1999), divided types of review in four categories.

- **Traditional narrative review**

A traditional narrative review is based on qualitative approach of studies. The experience and the subjectivity of the author is the most important thing. Usually traditional reviews are conducted by scientists who are experts in synthesizing research literature. However, it is well known that traditional narrative reviews suffer from the absence of a clear and objective method. It is vulnerable to unintentional bias selection and that leads to a number of methodological pitfalls. Also the emphasis is on a statistically significant result rather than on effect size (Lipsey & Wilson 1993). Of course carefully done studies, using the traditional narrative review exist and give a reliable result.

- **Meta-analysis from published data**

In this type of review researchers collecting data from published studies and calculate a total effect. However, neither this method promises perfect results. Because studies with negative results are usually not published, publication bias is again a serious problem. There also exists the major problem of heterogeneity. Even when studies have heterogeneity between them the research result will be published.

- **Meta-analysis with individual data**

Researchers in this type of meta-analysis, take the original data directly from the responsible scientist. Thomas, Sanyath & Benedett , (2014) have characterized the method of individual data as the gold standard of meta analysis. These data can be re-analyzed in a meta analysis and can produce more reliable results (Stewart & Tierney 2002.) This method has significantly less possibility of publication bias compared to the previous. However, it has some disadvantages as well. It requires a lot of time, effort and cost and it is quite difficult to change and improve the quality of data.

1.4.1 Narrative review vs. Meta-Analysis

A lot of questions have been raised related to what meta-analysis has to offer that narrative review cannot do.

For a topic usually there is huge number of studies. Although each study may try to answer the same questions there are many differences such as variables and size of samples. The results of each study will be compared and should lead to a conclusion. Traditional narrative reviews have significant disadvantages against the method of meta analysis

Narrative review, the traditional approach of research, is based on the selection of the data available and making comments about a collection of studies. The estimates are qualitative, not quantitative, and do not include the calculation of effect sizes that examine, the strength (or lack) of the effectiveness of an intervention. While narrative review can be useful in some cases, meta-analysis is a superior technique for integrating the literature on a topic.

It is generally accepted that traditional narrative review leaves room for subjectivity. It is common for two researchers to take two different results from the same body of literature. Meta-analysis seems to be a more objective method of research.

Although narrative reviews may be carefully done and lead to an extensive overview they are sensitive to publication bias (Blettner, Sauerbrei, Schlehofer, Scheuchenpflug & Friedenreich,1999). Apart from publication bias, the majority of the studies suffer from heterogeneity

The methodology followed by traditional review is no longer trustworthy, lacks transparency and replicability and it is prone to bias and erroneousness. A traditional review focuses on statistical significance which depends heavily on sample size and also a null finding does not carry same weight as a significant finding. Based on this method, researchers may draw different conclusions. For these reasons, a statistical approach is much more appropriate nowadays as it consist a more formal process of research.

1.5 Advantages and criticism of meta-analysis

1.5.1 Advantages of meta-analysis

There is no doubt that meta-analysis helps scientists to conduct more accurate studies and research. This chapter focuses on the advantages of this method. The most important of them are presented below (Pinaretos & Xekalaki, 2000, Dickersin & Berlin, 1992)

- It is a structured process which provides an objective synthesis of many relevant studies in order to reach a unique final result.
- It is possible to process large number of studies, data and information, larger than the traditional methods and represent findings in a more accurate manner.
- With meta-analysis a researcher is able to identify relationship between the studies. This is not always feasible in other methods of research.
- Gives the researcher the opportunity to weight differently each study based on the sample size each of them had used.
- This methodology allows a bias evaluation of the existing studies by calculating the number of studies that led to non-significant effect.
- Meta-analysis disposes a higher statistical power in detecting a significant effect of the single studies.
- The combination of many primary studies provides the opportunity to avoid bias that single studies may include.
- Sometimes it can work as a publication bias detector
- Researchers can calculate the effect size where the alternative hypothesis is different from the null hypothesis in a way that can determine if the importance of a finding is statistically significant

1.5.2 Criticism of Meta-Analysis

Meta-analysis has been widely welcomed as a statistical tool and as a scientific approach to combining and analyzing studies. But, it is a fact that meta-analysis is not the solution for every research problem and through the years, despite the great acceptance, intense criticism has been observed. It is crucial that potential researchers are informed of these problems. Critics present multiple arguments against meta-analysis. The most important of them are presented below (Thompson & Pocock, 1991, Dickersin & Berlin, 1992).

- The first problem that a researcher will meet is the "mixing apples and oranges problem". Comparisons between studies that include different subjects, variables or different measure techniques are unsuitable. The result may be affected in a wrong way by incomparable data and the summary effect may not take into consideration important differences across studies. It is a fact that a researcher cannot find studies with the same characteristics so inevitably studies will differ. It is up to the meta-analyst which studies should be included and it is sure that the opinions will vary from scientist to scientist. The solution to this problem is rather empirical. The researcher should check if the results are statistically affected by the characteristics of each study.

- Another major problem is that the quality of a meta-analysis is highly dependent on the studies that are taking part. This problem is often called garbage in, garbage out,(Eysenck, 1978, Hunt 1997) and suggests that if in a meta-analysis many low quality studies are taking part, the basic errors will affect the meta analytic result. Poorly designed studies are included along with good ones. This could mean that only high quality studies should be taken under consideration. This is called "best evidence synthesis". The magnitude of the effect is unrelated to the value of the study. Even if there is no significant difference in effect studies, there may be a significant difference in variance between the studies. This problem can also be solved empirically by investigate the differentiations of the results between poor and good studies. Some meta analysts suggests that poor quality studies could be included in a meta analysis under the condition that a predictor variable should be added that reflects the quality of primary studies and the effect of quality in the effect size. Also, many scientists believe that when researchers choose themselves the studies that are taking part, meta-analysis is highly dependent on the researcher's opinion and inevitably it is not easy to avoid subjectivity, which destroys the usefulness of meta-analysis.

- One more major problem of meta-analysis is publication bias. Studies with significant findings are more likely to be published than studies with non-significant findings. One solution is to examine the results of dissertations,

unpublished papers and abstracts of meetings. (Publication bias will be presented in detail in following chapter).

- Some critics say that another vulnerable spot of meta-analysis is that it can lead to simpler results because it draw attention to the whole effects rather than takes into consideration other factors.
- Another negative result is that the combination of many non-statistical significant studies can lead to a statistical significant result. This is known as Lindley' s paradox and clarify that a random effect can prove to be a significant factor when the observations number is big enough (Pinaretos & Xekalaki, 2000).
- To conduct a meta-analysis requires a lot of time and effort and this does not count as an asset of the method.

1.6 Publication Bias

It is practically impossible to know how many studies exists for a specific topic of research. Most researchers rely on published studies to conduct a meta-analysis. Publication bias exists if the probability that a study will be published depends on the result of the study (Scargle, 2000) and when the studies included in a meta-analysis differ from all the studies that should have been included. It is a problem that affects the researchers who do not only work on meta-analysis but also affects those who works in narrative review as well. But it is a fact that research in meta-analysis is more interested in this phenomenon. The ideal scenario is to include all the studies but this is quite rare even with the development of electronic searching.

Sterling (1959) was the first to find the existence of publication bias. He reviewed four psychology journals in one year (1955-1956). He found that 97% of the studies rejected the null hypothesis (Dickersin, 1990). The impact of publication bias in meta-analysis is the reason for the great interest in this subject. The first reported publication bias in a clinical trial meta-analysis was by John Simes (1986). He realized that only when registered trials were included then the treatments being compared did not differ significantly, But when only published trials were included, a statistical significant result was found. This conclusion suggested a relation between statistical significant results and publication bias (Dickersin, 1997)

It has been observed that usually studies with larger than average effect size are more likely to be published. And the studies with larger effect size are usually statistical significant (Dickersin, 2005). That means that the publication of a study is related to the result. There is a trend to publish positive rather than negative results. The result of this trend is to make it difficult for researchers to interpret the meta-analytic results. Easterbrook, Berlin, Gopalan & Matthews (1991) tried to find studies and follows them to discover which of them were published through the years (Borenstein, Hedges & Rothstein 2007). The result is that statistically significant studies (61%-86%) are more likely to published than the others and when finally published they observed a delay.

Rothstein (2006) made an attempt to explore if unpublished studies are included in meta analysis in Psychological Bulletin between 1995 – 2005. She found that almost 30% did not include unpublished data. Clarke & Clarke (2000) explored the health protocols of Cochrane Library, studied in detail references and review and found that 92% of the references are scientific published articles, 4% conference proceedings, 2% unpublished work, 1% were book chapters. Some researchers like Weisz (1995) suggested unpublished data should not be included because the quality of the study is affected negatively by them.

Of course scientists through the years have identified some other sources of publication bias. Some examples are (Borenstein, Hedges & Rothstein, 2007):

- Language bias (Egger 1997, Juni 2002). They say that literature and studies in English language are more likely to be searched.
- Availability bias: Studies and journals that are accessible more easily are most likely to be included.
- Cost bias: Expensive journals and papers are less likely to be selected rather than free low cost ones.
- Citation bias (Gotzsche 1997, Ravnksov 1992, Carter 2006) : Studies with statistically significant results are cited better and are more likely to be included in a systematic review

1.6.1 Addressing publication bias

The existence of publication bias directly affects the outcome of a meta-analysis. Scientists have developed some ways to deal with this phenomenon in the best possible way.

- The most common method of visualizing publication bias is the funnel plot. Proposed by Light and Pillener (1984), the idea is that it displays the relationship between effect size and study size. If the effect sizes are plotted near the average, and studies with low precision will be spread evenly on both sides of the average and scatter around producing a funnel shape. When negative studies are not included and gaps in the plot are observed then it is an indication of publication bias.

  Usually in a funnel plot effect size is represented on the X axis and sample size or variance on the Y axis. Sometimes the appearance of the standard error or variance on the Y axis can be replaced by the standard error,

  Figure 2 shows an example of a funnel plot showing no publication bias. Each dot represents a study (e.g. measuring the effect of a certain drug) the y-axis represents study precision (e.g. standard error or number of experimental subjects) and the x-axis shows the study's result (e.g. the drug's measured average effect).

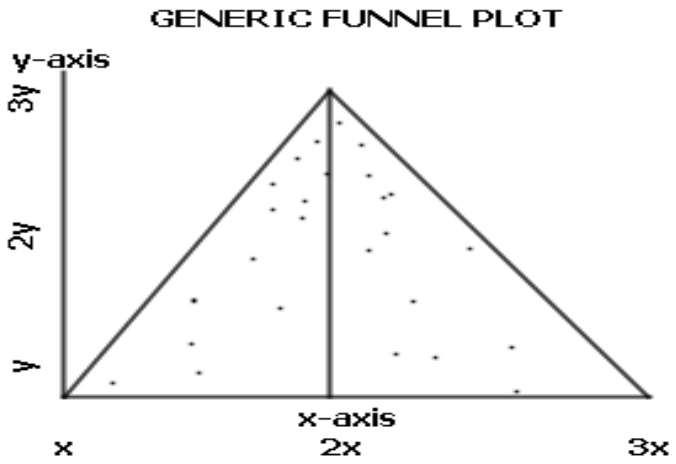

Figure 2: Presentation of a funnel plot

However, detecting publication bias through a funnel plot is a simple procedure but it can be misleading. If there is no symmetry in a funnel plot it can also be a sign of heterogeneity (Terrin, Schmid, Lau & Olkin , 2003). Heterogeneity will be analyzed in detail in following chapter). Another problem occurs when a meta-analysis includes a small number of studies. In this case it is difficult to explain the funnel plot.

Publication bias may lead the funnel plot to asymmetry. It is observed to be more small studies on the right side than on the left. Trying to correct this problem Duval and Tweedies introduced the "Trim and Fill" method. This method focuses on correcting the funnel plot asymmetry arising from publication bias. Remove the studies on the right side of the funnel plot (and in general the extreme studies) and compute again the effect size until the funnel plot is symmetric. Researchers can estimate the number of missing studies. However, this method does not take into consideration other reason for funnel plot asymmetry besides publication bias.

- Another way to confort the problem of publication bias is to calculate the number of studies that are confirming the null hypothesis. This method was presented by Rosenthal (1979) as File drawer analysis and is called File-Safe N by Cooper (1979). Rosenthal suggested the calculation of the number of the missing studies we need to unite, until p-value became non-significant. Hypothesize that mean effect in these studies is zero and focuses on statistical significant. If there is a need for a few studies to confirm the null hypothesis then the true effect was not zero. If we need a large number of studies to confirm the null hypothesis then there is no reason to be concerned (Borenstein, Hedges, Rothstein, 2007).

  This method uses the individual Z – score , ΣZ is the sum of the individual Z score and N is the number of studies (Wolf F.M.,1986).

$$N_{fs.05} = (\Sigma Z/1.645)^2 - N \text{ used to compute N for P=0.05}$$

$$N_{fs.01} = (\Sigma Z/2.33)^2 - N \text{ used to compute N for P=0.01}$$

This method has its disadvantages as well. The major issue is that this approach have proved inappropriate for a meta analysis focused on effect size (Borenstein, Hedges, Rothstein, 2007). Scargle (2000) has criticized Rosenthal's method, saying that he did not take into consideration the bias in the "file drawer" of unpublished studies, and thus can give confusing and misleading results.

- Orwin (1983) proposed File Safe N with a differentiation. The difference is that in this method meta-analysts can calculate how many of the missing studies would bring the total effect to a level other than zero.

1.7 Methods of Meta analysis

The first thing researchers look at when they conduct a meta-analysis is effect size. It encodes relationship of interest into a common index and standardized findings across studies that can be directly compared.

Standardized indexes such as standardized mean difference, correlation coefficient, odds-ratio, can be effect sizes as long as they meet the following requirements:

- Effect size should be comparable across studies
- Represent the magnitude and direction of the relationship of interest
- Is independent of sample size

When all the studies in a meta-analysis were equally precise we calculate the mean of the effect size. But some studies, especially studies that carry a biggest amount of information, are weighted more. This is the reason that in meta-analysis scientists computes the weighted mean rather than simple mean (Borenstein M., Hedges L., Rothstein Hannah, 2007). The calculation of the standard error is used to weight the larger studies.

Two models of effect size are used in meta-analysis, the fixed effect model and random effect model. These two methods differ regarding the assumptions about each study that can lead to different ways of assigning weights. The meta-analyst must choose, a priori, between a fixed or random effect model. The decision is based on this: Do meta-analysts expect studies to estimate a single population parameter? If the answer is yes the fixed effect model should be used. If the answer is no (and this is usually the case), the random effect model is the appropriate method.

1.7.1 Fixed effects model

In fixed effect models, researchers assume that there is one true population effect that all studies are estimating, share a common effect size μ and all of the variability between effect sizes is due to sampling error. The observed effects will be distributed around μ with a variance σ^2.

$$T_i = \mu + e_i$$

Where: T_i= the observed effect, μ = the true effect, e_i=the sampling error

As mentioned before, studies with more information are weighted. The weight for each study is the inverse variance.

This can be applied to log odds ratios and rarely to logistic regression. It is useful to mention that in Manzel – Haenszel (1959) method the odds ratios and not their logarithms are weighted inversely according to their variances (Spector, Thompson, 1991).

Then, the weighted mean is computed as the sum of effect size multiplied by the weight divided by the sum of the weights.

$$\overline{T}_{\cdot} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}$$

And the variance is the inverse sum of the weights.

$$V_{\cdot} = \frac{1}{\sum_{i=1}^{k} w_i}$$

The Standard error is computed as the square root of the variance.

$$SE(\overline{T}_{\cdot}) = \sqrt{V_{\cdot}}$$

1.7.2 Random Effects Model

When a meta-analyst decides which studies to include in a research, studies should have enough in common in order to be comparable. However that doesn't mean that these studies are identical. The effect size of each study is not the same in all studies. In this method meta-analysts assume that there is a distribution of true effect sizes. In this case researchers calculate the mea of the population of true effects. The differences in the observed effect sizes are due to real differences or to wrong sampling. In this method we can generalize the result. Random Effects model are preferred on methological grounds.

$$T_i = \theta_i + e_i = \mu + \varepsilon_i + e_i$$

Where: T_i: the observed effect θ_i is the true effect, determined by the mean of all true effects (μ) and error ε_i.

Under this model we add in another component the between group variance τ^2

As in the Fixed Effects Model, the weight assigned to each study is

$$w_i = \frac{1}{V_i}$$

To calculate the random effects weight we simply add τ^2 to the fixed effect variance.

The weighed mean is computed as the sum of the effect size multiplied by weight divided by the sum of the weights.

$$T_.^* = \frac{\sum_{i=1}^{k} w_i^* T_i}{\sum_{i=1}^{k} w_i^*}$$

The variance is calculated as the inverse sum of the weights

$$V^* = \frac{1}{\sum_{i=1}^{k} w_i^*}$$

The standard error is computed as the square root of the variance

$$SE(\overline{T_.^*}) = \sqrt{V_.^*}$$

1.8 Testing heterogeneity

The accuracy and the validity of a meta-analysis depend on how homogenous are individual studies with each other and whether all the effect sizes are estimating the same population, so that their results can be combined in order to calculate the total result. If homogeneity is rejected the distribution of effect sizes is heterogeneous. In order to achieve a good result it is essential to be homogenous in methodological design, in the populations under investigation. Of course, the results of the individual studies is expected to have a certain amount of volatility due to chance and of course some scientists strongly believe that, since clinical and methodological diversity always exist, statistical heterogeneity is inevitable (Higgins 2003) However, when the results of the individual studies included in the meta-analysis show greater heterogeneity than expected, the calculation of an overall result can lead to wrong conclusions. Of course, the studies whose results differ significantly from the results of the majority of studies should not be simply rejected because of this disagreement, but the various characteristics of the methodological design or data analysis that led to this discrepancy should be thoroughly examined.

Testing heterogeneity is a main concern of a researcher who conducts a meta-analysis. Heterogeneity tests the degree of dissimilarity in the results of individual studies (Walker et al., 2008).

When synthesizing a number of studies in one meta-analysis, it is believed that each study provides a sample estimate of the size of effect that represents population effect size. If a series of independent studies provide an homogenous - common estimate of the population effect size – it is more likely that studies included in meta-analysis are homogenous.

It is true that it is a rare phenomenon to include in a meta-analysis studies with similar protocols and it is more often that studies may be similar but with major differences such as different variation of a treatment. In this case, variation in study design is the main cause of incompatibility of the studies. When large differentiation exists between studies it is right to know if a meta-analysis is essential (Dickersin & Berlin, 1992)

Heterogeneity warns that it may not be suitable to include and combine all studies in one meta-analysis. In this case the meta-analyst may consider conducting separate meta- analytic synthesis (Wolf, 1986).

Formal methods to identify heterogeneity have low statistical power. This is the reason why scientists have developed informal methods to test heterogeneity. Examples of the methods used are graphical methods such as plots and compare studies results (Blettner M. et al., 1999)

Any variation can be manifested as heterogeneity.

- Clinical heterogeneity. Differences between treatment methods, patient characteristics, statistical analysis.
- Metrological differences
- Statistical heterogeneity: Differences in estimated indexes

The assessment of the heterogeneity of the results of studies to be included in meta-analysis can be performed either with forest plot, L'Abbé plot or by the application of appropriate statistical tests

Forest plot

This graphical test shows each study's Effect Size and 95% confidence interval in one table. We can observe the distribution of Effect sizes and determine if there is a variation between studies.

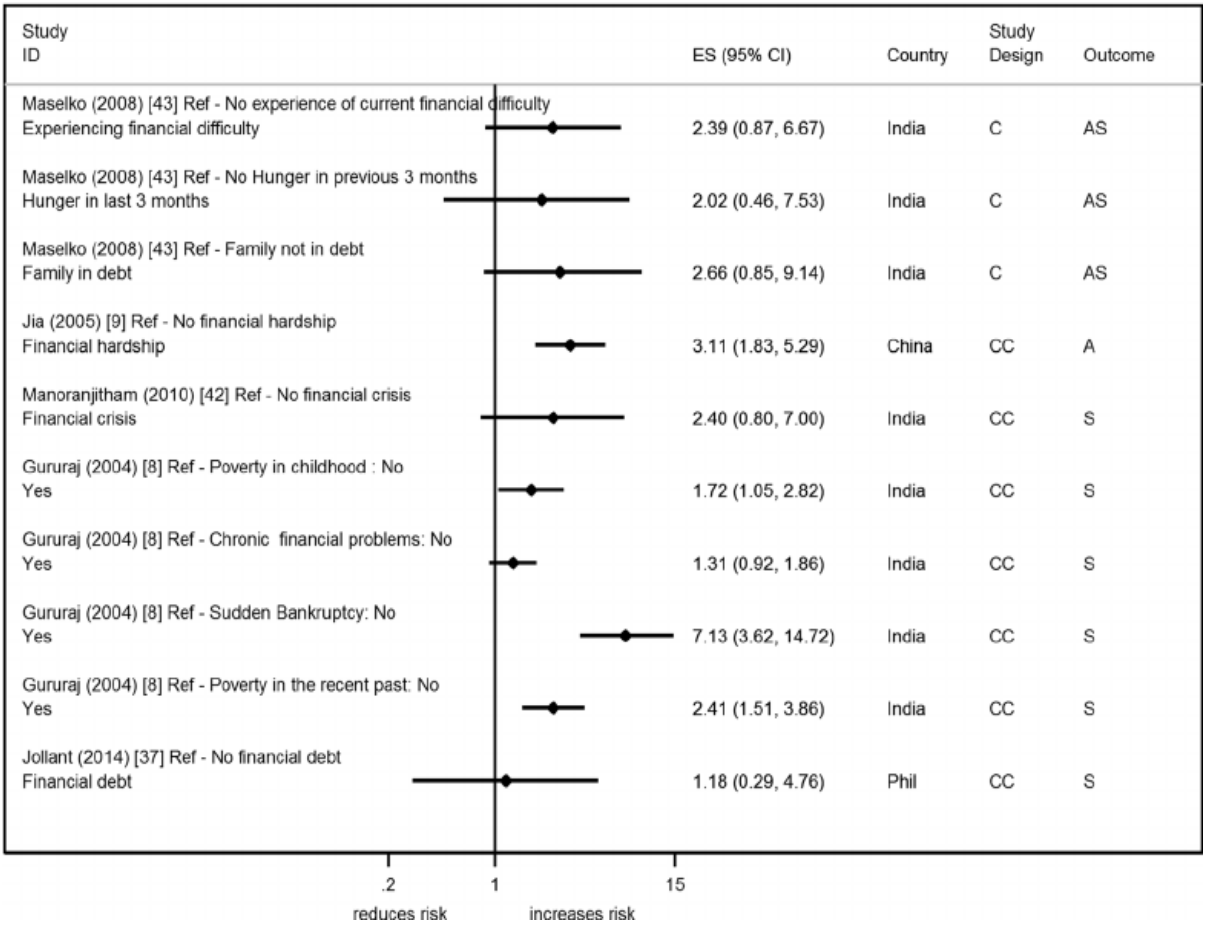

Figure 3: Example of Forest plot of studies reporting on financial difficulty and suicide/attempted suicide risk

Source: Duleeka W Knipe et al, 2015

L'Abbe Plot

L' Abbe Plot is like a simple scatter plot in which scientists can have a qualitative view of the data. The vertical axis represents the frequency for those belonging to the target category of under investigation determinant, while the horizontal axis represents the frequency for those referred as identifiers.(Galanis, 2008)

If the experimental determinant is better than the control the point will lie in the upper left of the l' Abbe plot between the vertical axis and the line. In the opposite case the point will be in the lower right of the plot between the horizontal axis and the line of equality.

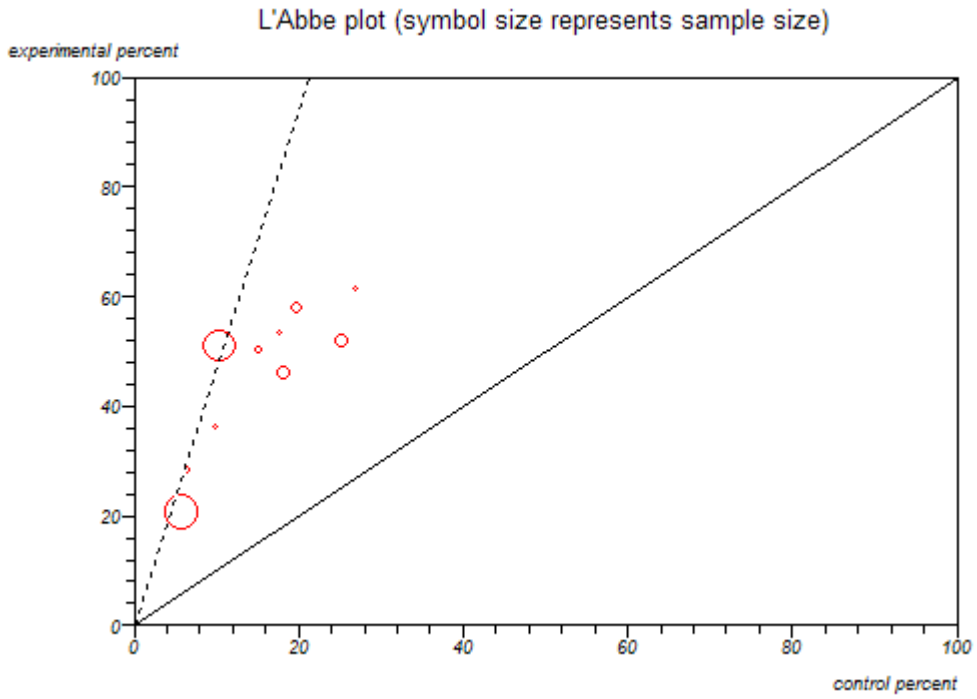

Figure 4: Exampe of L' Abbe Plot
Source: https://www.statsdirect.com/help/meta_analysis/labbe.htm

It is a fact that the use of the L' abbe plot is growing. Certainly it has several benefits. First of all the easy visual inspection of the level of agreement among trials.

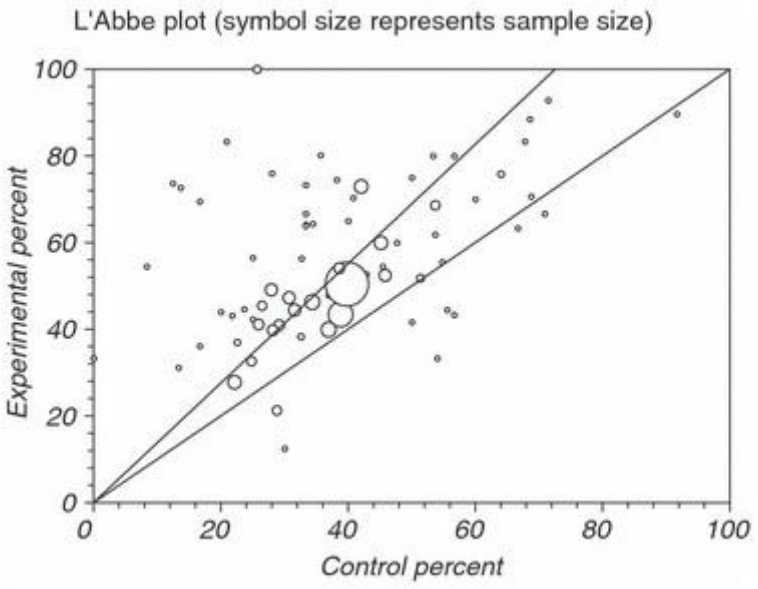

Figure 5: L'Abbe plot of placebo response rates (The lower line represents equality between experimental treatment and control, with circles above this line representing trials where experimental treatment was superior to control). As an example of the use of L' abbe plot is the meta-analysis. (Ford & Muayyedi, 2010)

Source: Ford & Muayyedi, 2010

Q method

If researchers conduct multiple experiments and used the same design, the same instruments, the same program for the same length of time, but used a different group and number of people, would they get the same result time and time again? Everybody wants this to be case, however it is known that even replicating with precision, researchers know that at least some variation of the estimate effect size exists.

The traditional method to compute heterogeneity is the Q method, a statistical test defined by Cochran (1954). The Q statistic is simply a weighted sum-of-squares differences between individual study effects and the pooled effect across studies, with the weights being those used in the pooling method.

Q estimates the total observed study-to-study variation and tests whether observed variability is more than you would expect from sampling error.

Another factor that scientists should address is, if the variation is due to sampling error alone or to some other, yet unexplained, factors.

Under the hypothesis of homogeneity among the effect sizes, the Q statistic follows a chi-square distribution, with k – 1 degrees of freedom and k is the number of studies included in meta-analysis. If homogeneity is not rejected the hypothesis usually leads the meta-analyst to use a fixed-effects model because it is assumed that the estimated effect sizes only differ by sampling error. On the other hand, rejecting the homogeneity assumption can lead to the use of a random-effects model that includes both within- and between-studies variability.

Ho: Effect sizes are the same across all studies

$$H_0 : ES_1 = ES_2 = \cdots$$

Original form of Q equation proposed by Cochran (1954) and defined as follows by (Hedges & Olkin, 1985)

$$Q = \sum w_i (ES_i - ES)^2$$

Because Q is a standardized measure, the expected value depends on the degree of freedom and not on effect size under the assumption that all the studies share the same effect size.

This test should not be used to justify a model

Failure to reject the null does not mean there is no heterogeneity in ES, only that you do not have enough evidence to detect it in your sample

It is also a fact that Q has low statistical power when the number of studies are small (Gavaghan et al, 2000) and Higgings (2003) claims that Q is powerful when number of studies is large. Also, non-significant results for the Q test with a small number of studies can lead a reviewer to erroneously use a fixed-effects model when there is true heterogeneity among the studies and vice versa. From Dickersin & Berlin (1992) is known that heterogeneity is statistical significant when $p < 0.10$.

Another tool for calculating heterogeneity is τ^2, also called the between study variance. τ^2 heterogeneity is controlled by Q method as part of the random effects meta-analysis. As the τ^2 depends on a particular effect metric used in a meta-analysis, it is not possible to compare the τ^2.

I^2 Method

Testing of heterogeneity is irrelevant to the choice of analysis. One alternative methods that scientists have developed to quantify the inconsistency between studies is I^2 and is computed as follows

$$I^2 = \left(\frac{Q - df}{Q}\right) * 100\%$$

Where: Q is the chi-squared statistic and df is its degrees of freedom (Higgins 2002, Higgins 2003). This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error.

For example, a meta-analysis with $I^2 = 0$ means that all variability in effect size estimates is due to sampling error within studies, assuming that the existence of heterogeneity could not be accepted. On the other hand, a meta-analysis with $I^2 = 50$ means that half of the total variability among effect sizes is caused by heterogeneity between studies (Medina T.H. et al, 2006).

Thresholds for the interpretation of I^2 can be misleading, since the importance of inconsistency depends on several factors. A rough guide to interpretation is as follows (Higgins & Thompson, 2002):

- 0% to 40%: might not be important
- 30% to 60%: may represent moderate heterogeneity
- 50% to 90%: may represent substantial heterogeneity
- 75% to 100%: considerable heterogeneity.

The importance of the observed value of I^2 depends on

- Magnitude and direction of effects and

- Strength of evidence for heterogeneity (e.g. P value from the chi-squared test, or a confidence interval for I^2)[1].

Advantages of I^2

Researchers strongly believe that I^2 is a much more appropriate and aplicable method addressing heterogeneity. Some advantages of the method are (Higgins, 2003):

- Pays attention to the effect of any heterogeneity on the meta-analysis
- Interpretation is intuitive—the percentage of total variation across studies due to heterogeneity
- Can be accompanied by an uncertainty interval
- Simple to calculate and can usually be derived from published meta-analyses
- Does not inherently depend on the number of studies in the meta-analysis
- May be interpreted similarly irrespective of the type of outcome data (eg dichotomous, quantitative, or time to event) and choice of effect measure (eg odds ratio or hazard ratio)
- Wide range of applications

Also I^2 method does not depend on the number of studies included in a meta-analysis.

The I^2 and τ^2 are directly related: the higher the τ^2, the higher the I^2 index. However, an advantage of the I^2 in respect to τ^2 is that I^2 indices obtained from meta-analyses with different numbers of studies and different effect metrics are directly comparable (Higgins & Thompson, 2002).

Choosing between Q and I^2

The Q method is much more useful when there is little variation between studies

When the results from these two methods coincide then there is no significant heterogeneity. The Q statistic is only useful for testing the existence of heterogeneity, but not the extent of heterogeneity. The I^2 index quantifies the magnitude of such heterogeneity and, if a confidence interval is calculated for it, then it can also be used for testing the heterogeneity hypothesis. the I^2 index with its confidence interval can substitute for the Q statistic, because it offers more information (Medina, 2006).

In case there is significant heterogeneity scientists have some options

- They avoid conducting a meta-analysis
- They usually apply random effects model
- They categorize studies into subgroups
- They apply a meta-regression method

[1] http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm)

1.9 Meta – regression

In single studies researchers use regression or multiple regression to examine and evaluate the relationship between moderators and the dependent variable. In meta-analysis researchers combine results in order to achieve a total effect size and this effect size is the dependent variable. According to Stanley & Jarrell (1989) meta-regression analysis is the statistical analysis of previously reported regression results. Meta-regression is a tool used to examine predictor variables on study effect size by using regression techniques and relate the effect size to the characteristics of the studies involved and examines the extent to which heterogeneity between studies is related to one or more study characteristics. Is an analogous technique to regression analysis using effect sizes as outcome and information extracted from studies as predictors. It also needs to account for weighting and choose between fixed or random effect model.

As shown by Higgins and Thompson (2004) "in contrast to simple meta-analysis, meta-regression aims to relate the size of effect to one or more characteristics of the studies involved".

Meta-regression is now used in medical, social and economic sciences. Some examples of the wide use of this method is the productivity spillovers on multinational companies (Gorg & Strobl 2001) or the business and price elasticities (Stanley & Doucouliagos 2009) or meta regression analysis on water policy cost savings (Bel, Fageda & Warner 2010). The use of meta-regression is rising with now over 100 new each year.

It is a quantitative method of conducting literature reviews. Scientists distinguised in three types of model.

- Simple regression
- fixed effect meta-regression
- random effect meta-regression

Cautions concerning meta-regression

Addressing heterogeneity sometimes leads to false (usually positive) results. It is crucial to recognize the possible disadvantages and pitfalls meta-regression methods may have. Some of them are analyzed below (Thompson & Higgins, 2002).

- False positive conclusions may be extracted when there is a small trials in a meta-analysis but there are many characteristics that explain heterogeneity in studies. Researches should ensure adequate sample sizes for moderation analysis
- It is not a rare unusual for meta-analysis to sometimes lead to false positive results. This is a dangerous phenomenon. The problem is that researchers are not able to check further the conclusions and it should be investigated as a new trial data set by multiple analyses.
- Researchers should specify and select a priori which predictors they are going to use in order to protect the research from false results. This is not something easy as most meta-analysis does not information and protocols publicly available.
- Furthermore, it is essential to set a limit to the number of covariates that a researcher include in a meta-analysis.
- Ensure scientific rationale for each moderator
- Beware the ecological fallacy (aggregation bias)
- Beware confounding with other study characteristics
- Because some meta-analyses contain a small number of studies it is a fact that the degrees of freedom may be small.
- Covariates tend to be collinear.

Meta Regression - Fixed Effects Model

Meta-regression fixed effects model assumes that (Borenstein M. et al, 2009):

$$y_i \sim N(\theta_i, v_i) \;\; \theta_i = X_t^t \beta$$

Where θ is the unknown true effect in the study, v is the variance for the estimated effect, x is the study level of moderating variable and β is a vector of coefficients for each of the moderating variables.

Meta Regression – Random Effects Model

The random effects model allows for the fact that there may be residual, unexplained variance in true effects across different studies. These residual variances are assumed to follow a normal distribution so that (Borenstein M. et al, 2009) :

$$y_i \sim N(\theta_i, v_i) \quad \theta_i = N(X_1^t \beta, \tau^2)$$

Where τ^2 is the residual variance in true effects. Programs to fit random effects meta-regression models using the method of restricted maximum likelihood (REML) have

been written in Stata (Sharp, 1998), SAS (van Houwelingen et al., 2002) and R (Viechtbauer, 2010).

Meta regression approaches (Morton, Adams, Suttorp MJ, et al, 2004)

1. The first approach is fixed effect meta-regression.
2. The second is random effect meta-regression
3. The third is control rate meta-regression
4. Fourth approach is Bayesian modeling.

A meta-regression can be a linear or logistic regression model. The majority of meta-regression analysis use as unit a study. For example the outcome of a study unit may be the log odds ratio.

PART 2: CASE STUDY-META ANALYSIS IN CAUSAL RELATION BETWEEN ENERGY CONSUMPTION AND ECONOMIC GROWTH

2.1 Introduction

In the practical part of this dissertation we are going to analyze and investigate the relationship between Gross Domestic Product (GDP) and energy consumption such as coal, oil, natural gas and total energy consumption. As Alfred Marshal said, GDP is the "outcome of the activities undertaken by the factors of production using other natural resources in the production sources during a given time period". The greater the activities efficiently undertaken the higher the national income that will result (Kuznets et al., 1941).Governments, scientists and in general members of the academic world are interested in finding out the impact of energy consumption in the economy.

Over the last decades it has been observed that the world economy has more than tripled in size. The importance of energy consumption on economic growth of countries all over the world has been acknowledged in the last fifty years because of the increasing concern over the lack of energy resources, especially after the two energy crises in 1973 and 1979. Many scientists firmly believe that the economic growth the world has known is responsible for the reduction of natural resources. Forecasts suggest that in fifty years as the world population accelerates we will face a massive reduction in resources. While GDP of leading economies continues to rise, economic growth in developing countries is even higher, with an average growth of 5.9% in total or 3.6% of per capita growth in low income countries, and 1.3% total or 0.8% per capita growth in high income countries (World Development Indicators, 2012). As well, world energy consumption has hugely increased in the 20th century and this increase will continue in the 21st century.

In the last decade many high quality analyses appeared to examine widely the relation between energy consumption and GDP. It is a fact that all major financial institutions like International Monetary Fund (IMF) or World Bank (WB) are constructing microeconomic and macroeconomic models in order to measure the impact of the price changes in oil and energy in general on economic growth. It is more than ever crucial to understand the relationship between growth and energy consumption in order to help countries and institutions to develop policies aimed to ameliorate in the reduction of the energy resources and climate change in general.

2.2 Literature Review

The relationship between energy consumption and economic growth analysing the direction between these two variables was first researched by Engle & Granger (1987). Many later studies claimed that the causal relation runs from economic growth towards increasing of energy consumption, as well as the increasing of energy consumption can lead to economic growth. On the other hand, some authors have shown that causal relationship among these two variables can be bi-directional, thus economic growth simultaneously effects energy consumption, and vice versa. The positive influence of energy consumption on economic growth can be seen in the positive externalities of energy, especially the impact of electricity on economic growth. The increase of electricity consumption is related to positive influence on health (e.g. through increasing usage of fridges) and education (radio, television...) contributing to economic growth and increase of the level of development.

For example, from authors who studied the relationship the relationship between energy consumption and GDP we have different and various results. Kraft & Kraft (1978), maybe the first study trying to find a relationship, found unidirectional causality from GDP to energy consumption in USA for the years 1947-1974. Two years later Akarca & Long (1980) said that the Kraft & Kraft (1978) results are false, The reason is that they do not find causality in this time period by two tear intervals. Erol & Yu (1987) studied several countries using Granger causality test and found a two way causality between energy consumption and GDP in Japan and Italy, one way causality in Eastern Germany and neutral relationship in France, UK and Canada. Hwang & Gum (1992) concluded that there is a two way relationship between them in Taiwan. More recently, Fatai (2002), studied these two variables in New Zealand found that there in no relationship between them. Al – Iriani (2005) for the countries that compose the Gulf Cooperation Council (Saudi Arabia, Oman, Kuwait, Bahrain, United Arab Emirates and Qatar), found that there is an one way causality from GDP to energy consumption. Rakhshan (2009) investigate this relation for Canada, China, Japan, Iran, and Russia.

Through the literature review we observe that there is a causality debate between energy consumption and GDP.

- One way causality running from GDP to Energy consumption. This hypothesis means that implementing policies that are aimed to increase GDP will lead to higher energy consumption
- One way causality running from Energy consumption to GDP. According to this hypothesis an increase in energy demand and consumption will lead to an increase in economic growth. On the other hand, energy conservation policies will slow economic growth.
- Two way causality between Energy consumption and GDP. This hypothesis implies that energy consumption causes economic growth and at the same time economic growth increases energy consumption
- Neutral Hypothesis. That means that there is no causal relation between energy consumption and GDP.

2.3 Data Collection

In this dissertation we are going to conduct a meta-regression in order to study if there is a connection between energy consumption and GDP. We have used the same data and the same philososphy as the Kalimeris et al (2014) study. The difference is that here we have used different dependent variables. We take into account 158 studies. It is important to say that initially the studies were 172 and Kalimeris et al. (2014) excluded some of them. Chronologically the examined studies cover the period from 1978 to 2011.

2.3.1 Variables

The dependent variable

The dependent variable is the causality between energy consumption and GDP. In the first model the dependent variable is binary whether the study concludes that there is a connection between energy consumption and GDP or not. In this case we are using binary logistic regression to examine the existence of a connection. In the second model we create a multinomial dependent variable which examines if the causality runs from GDP to energy consumption, or from energy consumption to GDP, or if there is no causal relationship between them.

Independent variables

The independent variables that are included in both models are:

- The publication year of each study examined. This variable has been divided into three categories. First for the published studies until 1990 (Kraft & Kraft, 1978; Akarca & Long, 1980; Erol &Yu,1987; Nachane et al.,1988. Second for studies until 2000 (Abosedra & Baghestani,1991; Yu & Jin,1992; Cheng, 1995; Murray & Nan, 1996; Masih & Masih, 1998; Cheng, 1999). Third and last category from year 2000 and after (Stern, 2000; Aqeel & Butt, 2001; Lee, 2005; Chontanawat et al., 2006; Chen et al..2007; Shuyun & Donghu, 2011)
- The time period each study examines. This variable is divided in two categories. First in studies with a shorter time period less than 30 years, (Ebohon, 1996; Murray & Nan, 1996; Masih & Masih, 1997; Lee . 2005) In this category there is one study with significant short time period (Abosedra et al. ,2009). The second category contains studies with longer time period from 31 years and above. (Yu & Choi, 1985; Erol &Yu, 1987; Chontanawat et al.,2006; Mahadevan & Asafu6Adjaye, 2007; Narayan & Wong, 2009)
- The next variable in question is the classification based on the development of the countries each study examines. This variable is divided into four categories. In the first category we include the countries which are participating in G7. (some studies in this category are: Yu & Choi, 1985; Murray & Nan, 1996; Huang et al. ,2008). In two other categories we have countries that are participating in the OECD (Nachane et al.,1988; Fatai et al., 2002; Narayan & Prasad, 2008) and countries that are not participating (Belloumi, 2009; Narayan & Smyth,2009; Ozturk & Acaravci,2011). And one more category is the high development countries (Murray & Nan, 1996;Fatai et al.,2004; Lee & Chang, 2008).

- Methodology is another independent variable, divided into six categories which describe the wide variety of econometric methodology. The categories are named as follows: Sims and Engle-Granger (Yu & Choi, 1985; Nachane et al.,1988; Murray & Nan, 1996; Chontanawat et al.,2008), Johansen-Juselius (Masih & Masih, 1996; Masih & Masih, 1998; Soytas & Sari, 2003; Shiu & Lam, 2004; Soytas & Sarib, 2006; Zamani, 2007), Toda-Yamamoto causality (Fatai et al.,2004; Wolde6Rufael,2005; Payne, 2009; Zhang & Cheng 2009), Perdoni panel cointegration (Lee, 2005; Lee & Chang,2008; Apergis & Payne ,2009a), ARDL test (Mozumder & Marathe, 2007; Squalli,2007; Halicioglu, 2007; Akinlo,2008) and other methods used (Altinay & Karagol, 2005; Chiou-Wei et al., 2008; Narayan & Prasad, 2008; Hu & Lin (2008).
- Energy category, divided into three variables. The first is category production of electricity (Murray & Nan,1996;Aqeel & Butt, 2001; Wolde6Rufael, 2006; Narayan & Prasad, 2008;Chih Chang, 2010). The second category refers to energy per capita and other (for example: Nachane et al., 1988; Asafu6Adjaye, 2000; Fatai et al., 2004) . The third category is total energy consumption (Ang ,2008; Karanfil ,2008; Warr et al., 2010,Chih Chang et al. 2011)
- Energy measurement. This variable is divided into 3 different categories . First literature reviews that used BTU (Fatai et al.,2004; Mishra et al.,2009). Next category oil equivalent and other (Yang, 2000; Soytas & Sari,2003; Lee, 2005). Last category is electricity production (Murray & Nan, 1996; Yang. 2000; Wolde-Rufael, 2006).
- The last independent variable indicates if the estimated causality refers to a single country (Yu & Choi, 1985; Murray & Nan, 1996; Chontanawat et al.,2006) or a group of countries (Narayan and Smyth, 2007; Lee & Chang , 2007; Jinke et al., 2008; Apergis & Payne, 2009; Chih Chang et al., 2011).

2.4 Analyses

Economic variables are often categorical rather than interval scale. In many cases the dependent variable is categorical and we examine how this variable is connected with other categorical variables (independent variables). In this case we should not conduct a linear regression but instead a logistic regression analysis. Logistic regression may be considered as an approach, that is similar to that of multiple linear regression, but takes into account the fact that the dependent variable is either categorical or continuous. The logistic regression is powerful in its ability to estimate the individual effects of continuous or categorical independent variables on categorical dependent variables (Wright 1995). We use logistic regression analysis to identify the relationships between the dependent variable and the independent variables. To achieve this, in our study we create two different models. In the first model we are going to use binary logistic regression and in the second multinomial logistic regression.

Other tools we are going to use to access the results are:

Wald statistic

Similarly to the t-test in linear regression, Wald statistic show us if the coefficient b for the independent variable (predictors) is significantly different from zero. In this case we believe that the predictor contributes significantly to the models outcome equation. The Wald statistic is calculated as follows:

$$Wald = \frac{b}{SE_b}$$

Where b is the value of the regression coefficient and SE is its standard error(Field, 2005). The Wald test is an important tool in regression analyses but its use needs caution because when the regression coefficient is large the se tends to be inflated leading to an underestimated Wald test (Menard, Scott.1995).

Odds ratio

A crucial tool for interpreting logistic regression is “exp(b)”. It is an indicator of the change in odds for a unit change in the predictor. When the predictor variables are categorical (as in our study) “exp(b)" is much easier to explain.

$$Odds = {P(event)}\Big/{P(no\ event)}$$

The previous equation shows that the odds of an event is the probability of an event occurring divided by the probability of the event not occurring.(Field, 2005)
CI stands for confidence interval and this option requests the range of values that we are confident that each odds ratio lies within. The setting of 95% means that there is only a $p < .05$ that the value for the odds ratio, exp(B), lies outside the calculated range.

2.5 First model – Binary logistic regression and results

2.5.1 Theory of Binary Logistic Regression

Binary logistic regression estimates the probability of a characteristic, given the values of explanatory variables, in this case we use as an example a single categorical variable.

$$\pi_i = \Pr(Y_i = 1 | X_i = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

Let Y be a binary response variable and X = (X1, X2, ..., Xk) be a set of explanatory variables which can be discrete, continuous, or a combination. xi is the observed value of the explanatory variables for observation i.

Or, $$\log(\pi_i) = \log\left(\frac{\pi_\iota}{1-\pi_i}\right) = \beta_0 + \beta_1 x_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

2.5.2 Application

In our model, we perform binary logistic regression with the SPSS programme.

We first see from the Omnibus Test of Model Coefficients that the model is statistical significant.

		Chi-square	df	Sig.
Step 1	Step	73,428	16	,000
	Block	73,428	16	,000
	Model	73,428	16	,000

Table 1: Omnibus Tests of Model Coefficients

From the Hosmer and Lomeshow test, which evaluates the goodness-of-fit of the model we see that we have a good model. (p-value = 0,375 > 0)

Step	Chi-square	df	Sig.
1	8,620	8	,375

Table 2: Hosmer and Lemeshow Test

In Table 3 we can see how the variables perform in the binary logistic regression. We observe that the explanatory variables “econometric methodology” (P<0.001) and “country” (P=0.28) are the only statistical significant independent variables in our model. The other variables do not make a statistically significant contribution to our model.

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Methodology			28,959	5	,000			
	Sims & E-G	-1,195	,355	11,300	1	,001	,303	,151	,608
	Johansen-Juselius	,131	,362	,132	1	,716	1,140	,561	2,316
	Toda - Yamamoto	-2,590	1,082	5,729	1	,017	,075	,009	,626
	Perdoni	-,905	,505	3,213	1	,073	,404	,150	1,088
	ARDL	,321	,429	,558	1	,455	1,378	,594	3,197
	Single country	-,278	,588	,224	1	,636	,757	,239	2,395
	Period (<30)	,063	,332	,036	1	,849	1,065	,556	2,041
	Country			9,251	3	,026			
	G7	-,397	,317	1,572	1	,210	,672	,361	1,251
	OECD	-1,010	,374	7,273	1	,007	,364	,175	,759
	High Development	-,814	,320	6,452	1	,011	,443	,237	,830
	Energy			1,692	2	,429			
	Electricity	,095	,492	,037	1	,847	1,099	,419	2,882
	Energy per capita and other	-,344	,564	,372	1	,542	,709	,234	2,143
	Measurement			3,520	2	,172			
	Btu	-,847	,452	3,509	1	,061	,429	,177	1,040
	Oil equivalent and other	-,691	,595	1,346	1	,246	,501	,156	1,610
	Year			,606	2	,739			
	2.00	,384	,810	,225	1	,635	1,469	,300	7,185
	3.00	-,040	,824	,002	1	,961	,961	,191	4,828
	Constant	,421	,796	,279	1	,597	1,523		

a. Variable(s) entered on step 1: Methodology, Single, periodnew2, countrynew, energynew, measurementnew, yearnew.

Table 3: Variables in the Equation

In detail, we clearly see that in the categorical variable econometric methodology the "Sims & E-G" and "Toda – Yamamoto" methodologies compared to the reference "Other" are statistically significant, contributing to the model. On the other hand methodologies "Johansen-Juselius" and "Perdoni" compared to the reference "Other" do not statistically significantly contribute to the model. For the variable "Country" "G7" compared with "Non OECD" does not statistically significantly contribute to the model whereas the categories "OECD" and "High Development" compared with "Non OECD" are statistically significant contributors to the model

Next, we are attempting to perform again a binary logistic regression without the non-statistical significant independent variables and the results are appearing in Table 4.

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Methodology			34,608	5	,000			
	Sims & E-G	-,199	,304	,428	1	,513	,820	,452	1,487
	Johansen-Juselius	-1,335	,344	15,047	1	,000	,263	,134	,517
	Toda - Yamamoto	,104	,334	,096	1	,756	1,109	,576	2,137
	Pedroni	-3,082	1,044	8,711	1	,003	,046	,006	,355
	ARDL	-,843	,453	3,462	1	,063	,431	,177	1,046
	country			12,133	3	,007			
	G7	,675	,263	6,572	1	,010	1,964	1,172	3,291
	OECD	,532	,255	4,360	1	,037	1,702	1,033	2,803
	High Development	-,206	,300	,472	1	,492	,814	,452	1,465
	Constant	-,954	,305	9,818	1	,002	,385		

Table 4: Variables in the Equation

In the final results we see that the methodologies that are contributing statistically significantly to the model are “Johansen-Juselius” and “”Pedroni” compared with the reference “Other” and the methodologies that do not contribute statistically significant are “Sims & E-G” and “Toda – Yamamoto” ” compared with the reference categories “Other”. For the other independent variable we see that categories “G7”, “OECD” compared with “Non OECD” are statistically significant contributors to the model but“High Development” compared with “Non OECD” does not contribute statistically significantly to the model.

.

We are also double checking the results with a Pearson chi-square test for the independent variables “methodology” and “country”.
The Pearson Chi-Square test is used in order to see if there is a relationship between two categorical variables. The larger its value, the larger the difference between the data and the null hypothesis of the hypothesis of the two variables. It tests whether or not a statistically significant relationship exists between a dependent and an independent variable.

Crosstab

			4. Methodology						Total
			Sims	J-J	Toda	Pedroni	ADRL	Other	
causalitynew1	"E to GDp" "GDP to E" "Two-way"	Count	148	163	73	51	43	45	523
		% within 4. Methodology	71,8%	88,1%	65,8%	98,1%	82,7%	65,2%	77,5%
	No connection	Count	58	22	38	1	9	24	152
		% within 4. Methodology	28,2%	11,9%	34,2%	1,9%	17,3%	34,8%	22,5%
Total		Count	206	185	111	52	52	69	675
		% within 4. Methodology	100,0 %	100,0 %	100,0 %	100,0%	100,0%	100,0 %	100,0 %

Table 5: Chi-square test for causality and econometric methodology

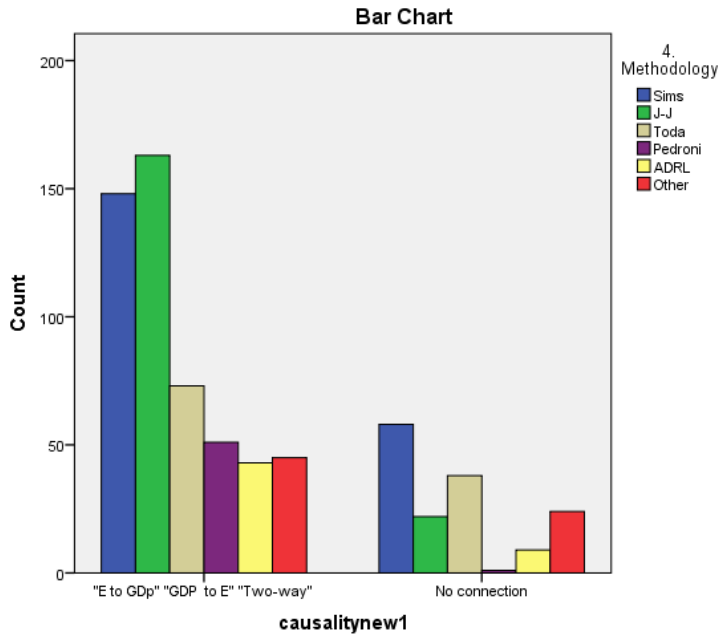

Figure 6: Bar chart of causality and econometric methodology

			countrynew				Total
			G7	OECD	High development	Non OECD	
causalitynew1	"E to GDp" "GDP to E" "Two-way"	Count	83	119	126	195	523
		% within countrynew	68,6%	73,0%	86,3%	79,6%	77,5%
	No connection	Count	38	44	20	50	152
		% within countrynew	31,4%	27,0%	13,7%	20,4%	22,5%
Total		Count	121	163	146	245	675
		% within countrynew	100,0 %	100,0%	100,0%	100,0%	100,0%

Table 6: Chi-square test for causality and country

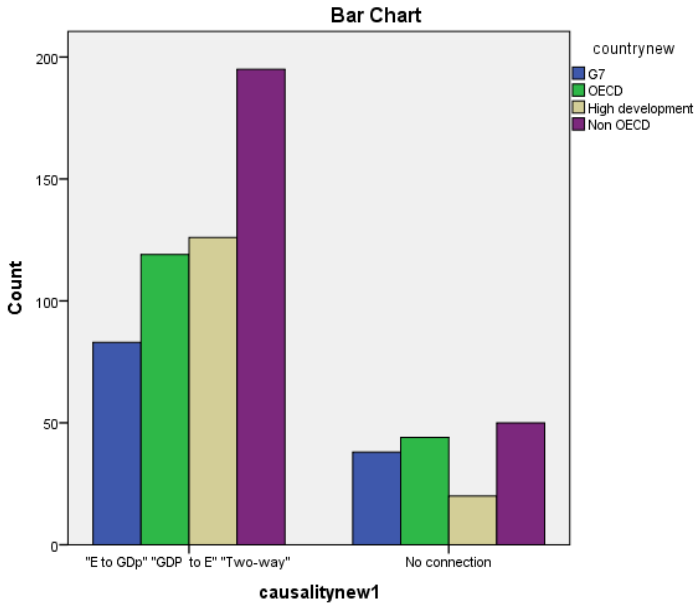

Figure 7: Bar chart of causality and country

So we can assume a strong relation between causality of GDP and Energy Consumption and econometric methodology as the p-value of those two is < 0.001. Again it is observed a strong relation between causality of GDP and Energy Consumption and country as the p-value is 0,002. In both cases we reject the hypothesis that the two variables are independent. So we can say that there is a connection between them and the appearance of causality in general, without knowing the direction of the causality.

2.6 Second Model – Multinomial Logistic regression and results

2.6.1 Theory of Multinomial Logistic regression

Multinomial Logistic regression is an expansion to logistic regression and estimates the association between a number of predictors and a multinomial outcome (Aldrich & Nelson 1984, Hosmer & Lemeshow 2000). The multinomial logistic regression model can be a useful tool for where the dependent variable is a discrete set of more than two choices (Agresti, 1996). Examples of such an outcome might be "yes", "no". "don't know". When estimating a multinomial model for a dependent variable with j categories, we estimate j-1 linear equations. Each model has its own intercept and regression coefficients and the predictors can affect each category differently. In a multinomial logistic regression model, the estimates for the parameter can be identified compared to a baseline category (Long, 1997).

Consider a dependent variable $Y_i = \ Y_1 \dots Y_k$ where k is the reference category

The model is given by the equation:

$$\ln \frac{P(Y_i=1)}{P(Y_k=1)} = g_i(x) = \ \beta_{i0} + \beta_{i1}x_1 + \beta_{i2}x_2 \dots, i = 1, \dots, k-1$$

$$\Rightarrow \ P(Y_i = 1) = P(Y_k = 1)\ e^{g_i(x)}$$

$$\Rightarrow \sum_{i=1}^{k} P(Y_i = 1) = 1 = P(Y_k = 1) + \sum_{j=1}^{k-1} P(Y_k = 1)\ e^{g_j(k)}$$

$$\Rightarrow P(Y_k = 1) = \frac{1}{1 + \sum_{j=1}^{k-1} e^{g_j(x)}} \Rightarrow \ P(Y_i = 1) = \frac{e^{g_i(x)}}{1 + \sum_{j=1}^{k-1} e^{g_i(x)}}$$

Multinomial Regression uses the maximum likelihood ratio to determine the probability of the categorical membership of the dependent variable. One of the reasons why Multinomial Logistic Regression is a good choice for this kind of data is that it does not assume normality, linearity, or homoscedasticity (Starkweather, 2011)

2.6.2 Application

When using multinomial logistic regression, one category of the dependent variable is selected as the reference category. We know that, parameters with significant negative coefficients decrease the likelihood of that response category with respect to the reference category. Parameters with positive coefficients increase the likelihood of that response category.

In our data, we are comparing the choices "GDP to E" and "Two-way" with "E to GDP". Our reference category is "E to GDP"

Firstly, we take a general look at the p-value of the independent variables. The tool that help us to see that is the Likelihood Ratio Tests. The test value used, to determine whether the independent variable has an effect on the dependent variable. The -2 log likelihood has a chi-square distribution, which can be used to determine whether the outcome of the test is significant. We want the p-value of Final to be <0.05. This criterion is confirmed by the results in the Table.

	Model Fitting Criteria	Likelihood Ratio Tests		
Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	557,071			
Final	473,668	83,403	32	,000

Table 7: Model Fitting Information

The Table below shows whether the model adequately fits the data. We want the p-values to be >0.05. In this case we could conclude that this model doesn't adequately fits the data.

	Chi-Square	df	Sig.
Pearson	317,639	226	,000
Deviance	330,895	226	,000

Table 8: Goodness-of-Fit

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	473,668[a]	,000	0	.
year	474,081	,413	4	,981
period	474,014	,347	2	,841
country	479,057	5,389	6	,495
Methodology	509,629	35,961	10	,000
measurement	489,185	15,518	4	,004
energy	484,205	10,537	4	,032
Single country	475,251	1,583	2	,453

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Table 9: Likelihood Ratio Tests

Now we want to know whether that effect is the same or not for each of the categories of the dependent variable. We observe that the only variables that have a significant overall effect on the outcome are “econometric methodology” , “measurement” and “energy”. We are attempting to do a new multinomial logistic regression with only the statistical significant independent variables .

.

The results appear in Table 10

causalitynew2[a]		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
GDP to E	Intercept	-,222	,548	,164	1	,686			
	Sims	-,204	,439	,216	1	,642	,816	,345	1,927
	J-J	-,706	,422	2,793	1	,095	,494	,216	1,130
	Toda	,238	,447	,284	1	,594	1,269	,528	3,048
	Pedroni	-1,046	,528	3,926	1	,048	,351	,125	,989
	ADRL	,312	,522	,358	1	,550	1,367	,491	3,802
	Other	0[b]	.	.	0	.	.	.	.
	Btu	1,525	,634	5,778	1	,016	4,595	1,325	15,936
	Oil	,046	,444	,011	1	,917	1,047	,438	2,501
	Other	0[b]	.	.	0	.	.	.	.
	Electricity	,174	,528	,109	1	,742	1,190	,423	3,350
	Energy per capita/other	,453	,300	2,276	1	,131	1,573	,873	2,834
	Total Energy	0[b]	.	.	0	.	.	.	.
Two-way	Intercept	-1,746	,611	8,157	1	,004			
	Sims	,382	,493	,602	1	,438	1,466	,558	3,849
	J-J	,593	,465	1,622	1	,203	1,809	,726	4,503
	Toda	-,899	,626	2,067	1	,151	,407	,119	1,386
	Pedroni	-,202	,551	,134	1	,714	,817	,277	2,408
	ADRL	,738	,591	1,562	1	,211	2,092	,657	6,656
	Other	0[b]	.	.	0	.	.	.	.
	Btu	,863	,742	1,353	1	,245	2,370	,554	10,147
	Oil	,978	,460	4,510	1	,034	2,658	1,078	6,551
	Other	0[b]	.	.	0	.	.	.	.
	Electricity	,959	,549	3,046	1	,081	2,608	,889	7,652
	Energy per capita/Other	,858	,292	8,615	1	,003	2,359	1,330	4,185
	Total Energy	0[b]	.	.	0	.	.	.	.

a. The reference category is: E to GDP.

b. This parameter is set to zero because it is redundant.

Table 10: Parameter Estimates

From Table 10 we observe that the majority of the variables have p-value greater than 0,05. That means that the adjusted odds ratio crosses over 1.0 and the association is non-significant.
If the p-value is < 0.05 and the adjusted odds ratio with its 95% CI is above 1.0, the risk of the outcome occurring increases that many more times versus the reference category outcome.

So we observe that in the first half of the Table has the outcome of GDP to E compared to E to GDP

- Pedroni compared to Other are less likely to contribute to GDP to E, 0.351 95% CI from 0.125 to 0.989
- Btu compared to Other is more likely to contribute to GDP to E.

in the second half of the Table has the outcome of Two way compared to E to GDP

- Oil compared to Other is more likely to contribute to E to GDP.
- Energy per capita and other compared to Total Energy are more likely to contribute to E to GDP.

2.7 Conclusions

Energy plays an important role in the economic growth of a country.
The purpose of this research is to determine the causal relationship between different measures of energy consumption and economic growth from 1978 to 2011. In this study we attempted to investigate if there is a causal relation between Energy Consumption and GDP and what kind of causality runs between them. For this purpose we used the meta-regression method and we constructed two different models. The techniques that were employed were at first binary logistic regression to determine if there is any kind of causal relation or not and second, a multinomial logistic regression to determine what kind of causal relation exists.

In the first model we conclude that the hypothesis that a causal relation exists, confirmed by the variables econometric methodology and country. In the second model, the statistically significant variables are econometric methodology, energy measurement and energy. We observe that the independent variable econometric methodology is the only common statistically significant variable in both models. However, the multinomial logistic regression model failed to show the exact causality direction of the model.

There is no doubt that further analyses should be made in the future with different data. Of course, apart from meta-analysis, single studies should be conducted in the future in order to add knowledge in this specific field of the science. For further investigation of the causal relationship between GDP and Energy Consumption, in order to completely understand it and produce more reliable results, is suggested the adding of other economic and environmental factors, such as energy prices, employment, and emissions of carbon dioxide. The study and usage of different independent variables could suggest improved, different or more specific policies for each country to implement.

Bibliography

Agresti, A. (1996). Categorical data analysis, New York: John Wiley & Sons.

Akarca, A. T., & Long, T. V. (1980). Relationship between energy and GNP: a reexamination. Journal of Energy Development, 5(2), 326-31

Aldrich, J. H., & Nelson, F. D. (1984).Linear probability, logit, and probit models, Sage.

Al-Iriani, M. A. (2005). Climate-related electricity demand-side management in oil-exporting countries—the case of the United Arab Emirates. Energy Policy, 33(18), 2350-2360.

Bel, G. & Warner, M. (2008). Does privatization of solid waste and water services reduce costs? A review of empirical studies. Resources, Conservation and Recycling, 52(12), 1337-1348.

Blettner, M. (1999). Traditional reviews, meta-analyses and pooled analyses in epidemiology. International Journal of Epidemiology, 28(1), 1-9.

Borenstein, M. (2009). Introduction to meta-analysis (1st ed.). Chichester, U.K.: John Wiley & Sons.

Cochran, W. G. (1954). The combination of estimates from different experiments. Biometrics, 10, 101-129.

Dickersin, K. & Berlin, J. (1992). Meta-analysis: State-of-the-Science. Epidemiologic Reviews, 14(1), 154-176.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. Journal of the American Medical Association, 263(10), 1385-1389.

Dickersin, K. (1997). How important is publication bias? A synthesis of available data. AIDS Education and Prevention, 9, 15-21.

Dickersin, K. (2005) Publication Bias: Recognizing the Problem, Understanding Its Origins and Scope, and Preventing Harm, in Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments (eds H. R. Rothstein, A. J. Sutton and M. Borenstein), John Wiley & Sons, Ltd, Chichester, UK.

Doucouliagos, H.& Stanley, T. D. (2009), Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis. British Journal of Industrial Relations, 47: 406–428.

Duval, S. & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics, 56(2), 455-463.

Easterbrook, P., Gopalan, R., Berlin, J., & Matthews, D. (1991). Publication bias in clinical research. The Lancet, 337(8746), 867-872

Egger, M. & Smith, G. (1997). Meta-analysis: Potentials and promise. British Medical Journal, 315(7119), 1371-1374.

Engle, R., & Granger, C. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. Econometrica, 55, 251-276

Erol, U., & Yu, E. S. (1987). On the causal relationship between energy and income for industrialized countries.Journal of Energy and Development, 13(1), 113-122.

Eysenck, H. (1978). An exercise in mega-silliness. American Psychologist, 33(5), 517-517.

Fatai, K., Oxley, L., & Scrimgeour, F. (2002, June). Energy consumption and employment in New Zealand: searching for causality. In NZAE conference, Wellington (Vol. 2).

Field, A. (2013).Discovering statistics using IBM SPSS statistics. Sage.

Fisher, M. R. (1944), Part-Time Employment of High School Students. Occupations: The Vocational Guidance Journal, 23: 161–163.

Ford, A. & Moayyedi, P. (2010). Meta-analysis: factors affecting placebo response rate in the irritable bowel syndrome. Alimentary Pharmacology & Therapeutics, 32(2), 144-158.

Gavaghan DJ, Moore AR, McQay HJ. An evaluation of homogeneity tests in meta-analysis in pain using simulations of patient data. Pain 2000;85:415-24.

Glass, G. (1976). Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5(10), 3-8

Glass, G., McGaw, B., & Smith, M. (1981). Meta-analysis in social research (1st ed.). Beverly Hills: Sage Publications.

Gorg, H. and Strobl, E. (2001), Multinational Companies and Productivity Spillovers: A Meta-Analysis. The Economic Journal, 111: 723–739.

Higgins, J. & Thompson, S. (2004). Controlling the risk of spurious findings from meta-regression. Statistics in Medicine, 23(11), 1663-1682.

Higgins, J. (2003). Measuring inconsistency in meta-analyses. BMJ, 327(7414), 557-560.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a metaanalysis. Statistics in Medicine, 21, 1539-1558

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. British Medical Journal, 327, 557-560

Hosmer Jr, D. W., & Lemeshow, S. & Sturdivant R.X (2013). Applied Logistic Regression, Second Edition, (Vol. 398), John Wiley & Sons

Huedo-Medina, T., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index?. Psychological Methods, 11(2), 193-206.

Hunt, M. (1997). How science takes stock (1st ed.). New York: Russell Sage Foundation.

Hunter, J., Schmidt, F., & Jackson, G. (1982). Meta-analysis (1st ed.). Beverly Hills [Calif.]: Sage.

Hwang, D., & Gum, B. (1992). The causal relationship between energy and GNP: the case of Taiwan. Journal of Energy and Development, 16(2), 219-26.

Juni, P. (2002). Direction and impact of language bias in meta-analyses of controlled trials: empirical study. International Journal of Epidemiology, 31(1), 115-123.

Kalimeris, P., Richardson, C., & Bithas, K. (2014). A meta-analysis investigation of the direction of the energy-GDP causal relationship: implications for the growth-degrowth dialogue. Journal of Cleaner Production, 67, 1-13.

Knipe, D., Carroll, R., Thomas, K., Pease, A., Gunnell, D., & Metcalfe, C. (2015). Association of socio-economic position and suicide/attempted suicide in low and middle income countries in South and South-East Asia – a systematic review. BMC Public Health, 15(1).

Kraft, J., & Kraft, A. (1978). Relationship between energy and GNP. J. Energy Development, 3(2), 401-03

Kuznets, S. (1941). National Income And Its Composition 1919 1938 Vol I., Publications of the National Bureau of Economic Research, Number 40.

Lipsey, M. & Wilson, D. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. American Psychologist, 48(12), 1181-1209.

Mantel N & Haenszel MW (1959). Statistical aspects of thee analysis of data from retrospective studies of disease. Journal of the National Cancer Institute ;22:719-48.

Menard, S ,(1995). Applied Logistic Regression Analysis. Sage Publications.Series: Quantitative Applications in the Social Sciences, No. 106.

Morton SC, Adams JL, Suttorp MJ, et al. Meta-regression Approaches: What, Why, When, and How? Rockville (MD): Agency for Healthcare Research and Quality (US); 2004 Mar. (Technical Reviews, No. 8.)

Orwin, R. (1983). A Fail-Safe N for Effect Size in Meta-Analysis. Journal of Educational Statistics, 8(2), 157.

Pearson K (1904). Report on certain enteric fever inoculation statistics. British Medical Journal 3:1243-1246.

Petiti D.B., (2000). Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine, Oxford University Press, USA,

Rakhshan S (2009). The Relationship between Energy Consumption and Iran's Economic Growth and Development with Special Reference to Oil, Natural Gas, and Electricity during 1984 – 2004. University of Pune India. A PhD Dissertation

Ravnskov, U. (1992). Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. British Medical Journal, 305(6844), 15-19.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86(3), 638-641

Scargle, J. D. (2000). Publication bias: The "File Drawer" problem in scientific inference. Journal of Scientific Exploration, 14, 91-106.

Sharp S. (1998) Meta-analysis regression. Stata Technical Bulletin 42(23):16 – 22

Simes, R. J. (1986). Publication bias: the case for an international registry of clinical trials. Journal of Clinical Oncology, 4(10), 1529-1541.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. American Psychologist, 32(9), 752-760.

Spector, T. D., & Thompson, S. G. (1991). The potential and limitations of meta-analysis. Journal of Epidemiology and Community Health, 45(2), 89–92.

Stanley, T. D. and Jarrell, S. B. (1989), Meta-regression analysis: a quantitative method of literature surveys. Journal of Economic Surveys, 3: 161–170.

Starkweather, J., and Moske, A. K. (2011). Multinomial Logistic Regression. www.unt.edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011.pdf

Sterling, T. (1959). Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance--Or Vice Versa. Journal of The American Statistical Association, 54(285), 30-34.

Stewart, L. & Tierney, J. (2002). To IPD or not to IPD?. Evaluation & The Health Professions, 25(1), 76-97.

Terrin, N., Schmid, C., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. Statistics In Medicine, 22(13), 2113-2126.

Thomas, D., Radji, S., & Benedetti, A. (2014). Systematic review of methods for individual patient data meta- analysis with binary outcomes. BMC Medical Research Methodology, 14(1).

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. Journal of Statistical Software, 36(3), 1-48.

Weisz, J., Weiss, B., Han, S., Granger, D., & Morton T,. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. Psychological Bulletin, 117(3), 450-468.

Wolf, F. (1986). Meta-analysis (1st ed.). Beverly Hills, Calif.: Sage Publications.

Πανάρετος, Ι., & Ξεκαλάκη, Ε. (2000). Εισαγωγή στη στατιστική σκέψη, Τόμος 2, Εισαγωγή στις πιθανότητες και στη στατιστική συμπερασματολογία. Εισαγωγή στις πιθανότητες και στη στατιστική συμπερασματολογία. Αθήνα