Crawford School of Public Policy

Is There Really Granger Causality Between Energy Use and Output?

Crawford School Working Paper No. 13 - 07

8 March 2013

Stephan B. Bruns
University of Jena and Max-Planck Institute of Economics
Jena, Germany

Christian Gross
Institute for Future Energy Consumer Needs and Behavior (FCN)
School of Business and Economics/E.ON Energy Research Center
RWTH Aachen University
52074 Aachen

David I. Stern
Crawford School of Public Policy
Australian National University
Canberra, ACT 0200
Australia

Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Abstract

We carry out a meta-analysis of the very large literature on Granger causality tests between energy use and economic output to determine if there is a genuine effect in this literature or whether the large number of apparently significant results is due to publication and misspecification bias. Our model extends the standard meta-regression model for detecting genuine effects using the statistical power trace in the presence of publication biases by controlling for the tendency to over-fit vector auto regression models in small samples. These over-fitted models have inflated type 1 errors. We find that models that include energy prices as a control variable find a genuine effect from output to energy use in the long-run. A genuine causal effect also seems apparent from energy to output when employment is controlled for and the Johansen procedure is used.

Keywords:

JEL Classification: Q43, C32, C52

Suggested Citation:

Bruns, Stephan B., Gross, C., Stern, David, I., 2013, *Is There Really Granger Causality Between Energy Use and Output*, Crawford School Working Paper No. 13 - 07, March 2013, Crawford School of Public Policy, The Australian National University, Canberra

Address for correspondences:

Stephan B. Bruns

s.b.bruns@googlemail.com

Christian Gross

gross@econ.mpg.de

David I. Stern

David.stern@anu.edu.au

Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

Is There Really Granger Causality Between Energy Use and Output?

Stephan B. Bruns^a, Christian Gross^b, and David I. Stern^c

^a University of Jena and Max-Planck Institute of Economics, Jena, GERMANY. E-mail: s.b.bruns@googlemail.com

^b Institute for Future Energy Consumer Needs and Behavior (FCN), School of Business and Economics/E.ON Energy Research Center, RWTH Aachen University, 52074 Aachen, GERMANY. E-mail: gross@econ.mpg.de

^c Corresponding author: Crawford School of Public Policy, Australian National University, Canberra, ACT 0200, AUSTRALIA. E-mail: david.stern@anu.edu.au

8 March 2013

Abstract

We carry out a meta-analysis of the very large literature on Granger causality tests between energy use and economic output to determine if there is a genuine effect in this literature or whether the large number of apparently significant results is due to publication and misspecification bias. Our model extends the standard meta-regression model for detecting genuine effects using the statistical power trace in the presence of publication biases by controlling for the tendency to over-fit vector autoregression models in small samples. These over-fitted models have inflated type 1 errors. We find that models that include energy prices as a control variable find a genuine effect from output to energy use in the long-run. A genuine causal effect also seems apparent from energy to output when employment is controlled for and the Johansen procedure is used.

JEL Codes: Q43, C32, C52

Acknowledgements: We thank Ariane Bretschneider, Maria Hennicke, Clemens Klix, Susanne Kochs, Natalija Kovalenko, Katja Mehlis, Stefanie Picard, Annemarie Strehl, and Silvia Volkmann for research assistance. We thank Paul Burke, Chris Doucouliagos, Kerstin Enflo, Ippei Fujiwara, Alessio Moneta, Tracy Wang, and participants at the MAER-Net Symposium in Perth in September 2012 for useful comments. David Stern acknowledges funding from the Australian Research Council under Discovery Project DP120101088. Stephan Bruns is grateful to the German Research Foundation (DFG) for financial support through the program DFG-GK-1411: “The Economics of Innovative Change”.

Introduction

Since 1978 (Kraft and Kraft, 1978), the literature on Granger causality between energy and economic output has grown rapidly and now consists of hundreds of papers. But despite attempts to review and organize this literature (e.g. Ozturk, 2010; Payne, 2010), the nature of the relationship between the variables remains unclear (Stern, 2011). It is important to understand these relationships because of the general role of energy in economic production and growth (Stern, 2011), the ongoing debate about the effect of energy price shocks on the economy (Hamilton, 2009), and the important role of energy in climate change policy. In this paper, we carry out a meta-analysis of the very large literature on Granger causality tests between energy use and economic output. Our goal is to determine whether genuine effects exist in this literature or whether the large number of apparently significant results is due to publication and misspecification bias.

The methods we use in this paper should also be applicable to other areas of research that use Granger causality testing. Granger causality techniques have been widely applied in many fields of economics including monetary policy (Lee and Yang, 2012), finance and economic development (Ang, 2008a), and energy economics (Ozturk, 2010) and also in other disciplines such as climate change (e.g. Kaufmann and Stern, 1997) and neuroscience (Bressler and Seth, 2011). But the results of Granger causality testing are frequently fragile and unstable across specifications (Lee and Yang, 2012; Ozturk, 2010; Stern, 2011). Meta-analysis is a method for aggregating the results of many individual empirical studies in order to increase statistical power and remove confounding effects (Stanley, 2001). Simple averaging of coefficients or test statistics across studies is, however, plagued by the effects of publication and misspecification biases. Publication bias is the tendency of authors and journals to preferentially publish statistically significant or theory-conforming results. In the worst-case scenario, there may be no real effect in the data and yet studies that find statistically significant results are published. This has led a prominent meta-analyst to claim that: “Most Published Research Findings Are False” (Ioannidis, 2005). In this paper, we show how meta-analysis can be used to test for genuine effects, publication, and misspecification biases in Granger-causality studies. Some of the techniques should also be useful in the meta-analysis of studies using other econometric methods.

We base our analysis on a fairly standard meta-regression model that controls for the effects of publication bias and exploits the statistical power trace to find genuine effects in empirical

literatures. This model regresses test statistics from individual studies on the square root of the degrees of freedom of each study. The slope coefficient then tests for the presence of a genuine effect in the literature and the intercept tests for the presence of publication bias. This is because when the genuine effect size is non-zero, an increase in the degrees of freedom implies an increase in the test statistics due to statistical power, whereas if there is no genuine effect the p-value of the test statistics will be uniformly distributed whatever the size of the sample. Granger causality tests present two challenges to the simple version of this model. The first is that the usual restriction test statistics have an F or chi-squared distribution and these must be converted to a common statistic with properties that are suitable for regression analysis. We transform the p-values of the test statistics to standard normal variates.¹ The standard normal distribution is also better than the commonly used t-distribution because the distribution is unaffected by degrees of freedom and we recommend its wider adoption in meta-analysis. The second challenge is the tendency for researchers to over-fit vector autoregression (VAR) models in small samples. These over-fitted models tend to result in over-rejection of the null hypothesis of Granger non-causality when it is false, especially in small samples. We control for these effects by including as a control variable the number of degrees of freedom lost in fitting the model.

A recent exploratory meta-analysis of 174 pairs of tests (each pair tests whether energy causes output and *vice versa*) from 39 studies uses a multinomial logit model to test the effect of some sample characteristics and methods used on the probability of finding Granger causality in each direction (Chen *et al.*, 2012). Chen *et al.* (2012) conclude that researchers are more likely to find that output causes energy in developing countries and that energy causes output in OPEC and Kyoto Annex 1 countries. Additionally, output is more likely to cause energy in larger countries and in studies with more recent data, but higher total energy use is likely to result in a finding that energy causes output. They also find that the standard Granger Causality test is more likely to find causality in some direction than are alternative methods. Though these findings are interesting, Chen *et al.*, (2012) do not address whether the causality tests represent a sample of valid statistical tests or are the possibly spurious outcomes of publication and misspecification bias. We test for whether there are actual genuine effects in this literature rather than just misspecification and publication selection biases. Additionally, we have a larger sample consisting of 574 pairs of causality tests from

---

¹ Stanley (2005b) similarly converts F and Chi-Square test statistics to normal variates.

72 studies selected from this vast literature of more than 400 papers. Our selection of papers is based on clearly defined and documented criteria.

The first part of our paper outlines our model for testing for genuine effects and publication and misspecification biases in the Granger causality literature. We then describe the choice of studies for our meta-analysis, followed by an exploratory analysis of the data. This includes a description of the data, a correlation analysis, and meta-significance tests. This analysis finds no genuine effect in the meta-sample as a whole but also shows the likelihood of severe misspecification biases. We then apply models that control for these misspecification biases to both the data as a whole and using dummy variables to various subsets of the literature. We find that there is still no genuine effect in the literature as a whole but find that models that include energy prices as a control variable have a genuine effect from output to energy use in the long-run. A genuine causal effect also seems apparent from energy to output when employment is controlled for and the Johansen procedure is used. This effect is more ambiguous because including capital weakens the effect and carrying out causality tests after imposing the cointegration restrictions is known to have inflated type 1 errors. It is possible that such a genuine effect is also evident in the sub-sample of all studies using macro-level variables for both energy and output. It is also possible that a genuine causal effect might be detected for some subset of countries or time periods but that is not tested in the present paper. The final section provides some suggestions and recommendations for future research.

Methods

Testing for Genuine Effects

Testing for the existence of a genuine effect in meta-data using meta-regression analysis is based on the idea that in the absence of publication and misspecification biases, and abstracting from genuine heterogeneity, the estimated effect size, $\hat{\beta}$, – in econometrics typically a regression coefficient(s) of interest – should have the same expected value across studies irrespective of their degrees of freedom, DF . But the precision, $\hat{\sigma}_{\beta}^{-1}$, of a consistent estimator of the effect size tends to increase linearly with the square root of the degrees of freedom as the parameter estimate converges in probability to the true value.² Therefore,

---

² Convergence is probabilistic and standard errors will vary across datasets in ways that are unrelated to sample size or degrees of freedom. Also, for cointegrated models the rate of

assuming for simplicity that the null hypothesis is $\beta = 0$, the related t-statistic should increase in absolute value linearly with the square root of the degrees of freedom if there is a genuine non-zero effect:

$$\frac{\hat{\beta}_i}{\hat{\sigma}_{\beta i}} = t_i = \alpha DF_i^{0.5} + u_i \quad (1)$$

$$u_i \sim t(DF)$$

where i indexes individual test statistics³ and α has the same sign as the genuine effect. The errors are thus predictably heteroskedastic as the variance of the t-distribution increases as the degrees of freedom decreases. This heteroskedasticity can be removed by converting the t-statistics to normal variates with the same p-values:

$$Z_i = \alpha DF_i^{0.5} + v_i \quad (2)$$

$$v_i \sim N(0,1)$$

It is usual to estimate a logarithmic version of (1) or (2), which Stanley (2005a, 2008) calls meta-significance testing or MST:

$$\ln|y_i| = \ln\alpha_0 + \alpha_1 \ln DF_i + \varepsilon_i \quad (3)$$

where y is the dependent variable from equations (1) or (2). Rejecting the null-hypothesis that $\alpha_1 = 0$ suggests that there is a genuine effect in the meta-sample. However, this functional form is undesirable. First, if we use t-statistics rather than normal variates, the heteroskedasticity of the t-statistics will introduce a spurious negative correlation between the test statistics and the degrees of freedom for low degrees of freedom once absolute values are taken. Therefore, normal variates are more appropriate. Second, due to taking absolute values and logarithms the error term will not have a normal distribution, and will also be heteroskedastic if there is a genuine effect. Though Stanley (2008) found (3) to be very powerful in large meta-samples of studies even in the presence of publication biases, this test suffers from inflation of type 1 errors (Stanley, 2008; Stanley and Doucouliagos, 2012).

---

convergence of the parameters of the cointegrating vector to their true value is faster – the super-consistency property – but we do not address this point in this paper.

³ Each underlying study often contains several model estimates and more than one test statistic may be computed with each model – for example tests of “short-run” and “long-run” causality.

The control of publication bias is an alternative motivation in the meta-regression literature. If journals will only publish, or authors only submit for publication, statistically significant results then, the larger the effect size must be the less the precision of estimation is in order to achieve a given p-value. If all results are equally likely to be accepted for publication there should be no relation between estimated effect size and the standard error. In the absence of publication bias, though the estimated effect size will tend to be closer to the genuine value the smaller the standard error but estimated effect sizes should be symmetrically distributed around the genuine value – the so-called funnel graph (Stanley, 2001). This suggests the following model:

$$\hat{\beta}_i = \gamma_0 + \gamma_1 \hat{\sigma}_{\beta i} + e_i \quad (4)$$

The test of $\gamma_1 = 0$, which Stanley (2005a) calls FAT (funnel asymmetry test) is a test for publication bias while γ_0 is an estimate of the value of the genuine effect adjusted for the publication bias. This relationship is exact when the genuine effect is zero (Stanley & Doucouliagos, 2011) and, therefore, is a suitable model for testing the null of no genuine effect.⁴ As (4) has heteroskedastic errors, Stanley (2005a) suggests that researchers divide both sides of (4) by the standard error and estimate the following model instead:

$$t_i = \gamma_0 \frac{1}{\hat{\sigma}_{\beta i}} + \gamma_1 + v_i \quad (5)$$

The same hypothesis tests apply to (5) as applied to (4) but it is now the intercept term which tests for publication bias and the slope coefficient is the estimate of the genuine effect. Stanley calls the test of $\gamma_0 = 0$ PET (Precision Effect Test). When we do not have information on standard errors, as in the case of most Granger causality tests, we can approximate the precision in (5) by the square root of degrees of freedom (Stanley, 2005b):

$$t_i = \gamma_0 cDF^{0.5} + \gamma_1 + \omega_i \quad (6)$$

---

⁴ However, it does tend to underestimate the absolute value of the genuine effect when it is non-zero because if the genuine effect is much larger than the standard error there is no need to select publications for significant effects. Only in smaller samples will there then be a linear relationship between effect size and standard error, while in large samples there will be no relation (Stanley and Doucouliagos, 2011). Therefore, publication bias should not actually be represented by a constant term. In any case, in the Granger causality literature we are not concerned with the size of the effect itself and so (4) is an adequate approximation. Stanley and Doucouliagos (2011) recommend to use the PET model (5) to test for a genuine effect and then use the PEESE model to estimate the size of the genuine effect if one exists.

where $c = E(\hat{\sigma}_{\beta i} / DF_i^{0.5})$. But (6) is simply (1) with the addition of a constant. So, PET can be motivated by the same statistical power argument as was used to motivate MST (Stanley and Doucouliagos, 2012). We, therefore, estimate (6) using normal variates:

$$Z_i = \alpha_0 + \alpha_1 DF_i^{0.5} + v_i \quad (7)$$

This model allows a neat decomposition of the sources of variance in the test statistics. By contrast, the intercept in the MST model (3) is a function of both the value of the genuine effect and publication bias. Granger causality test statistics are usually F or Chi-square distributed⁵ and in order to apply model (7) they need to be converted to normal variables. We convert them using the probit function - the inverse of the standard normal cumulative distribution. The transformation takes p-values of less than 0.5 and transforms them into negative normal variables with the significance levels for a one-sided hypothesis test. Values greater than 0.5 are transformed to positive normal variables with significance levels for a one-sided hypothesis test. For example, $probit(0.025) = -1.96 = -probit(0.975)$. To help intuition, we multiply these statistics by -1 so that more positive values are associated with rejecting the null hypothesis of no-causality at higher levels of significance. In the absence of publication bias, the intercept is expected to be zero $probit(0.5) = 0$. For these probit-transformed p-values the appropriate test for a genuine effect is a one sided test for a slope coefficient greater than zero. This is because test statistics equal to both zero and less than zero imply that there is no true effect.

We give equal weight to each test statistic from each paper and use heteroskedasticity robust clustered standard errors throughout. We estimate models separately for causality tests in each direction. There is little gain from joint estimation, as in most studies the degrees of freedom are the same for both tests. In our initial estimates, in addition to the preferred model (7) we also estimate (3) and a logarithmic version of (2) as a comparison.

Controlling for Misspecification Biases

The number of lags of the variables in a VAR is typically chosen using the Akaike Information Criterion (AIC) or other goodness of fit indicators. The AIC, in particular, tends to over-estimate the number of lags when degrees of freedom are low and also the VAR has a unit root or near unit root (Nickelsburg, 1985; Hacker & Hatemi-J, 2008). This problem is

---

⁵ Some of the test statistics in our study are actually t-statistics.

reduced in larger dimensional systems (Gonzalo and Pitarakis, 2003).⁶ Zapata and Rambaldi (1997) show that three different causality tests over-reject the genuine null of non-causality in small samples especially when there is over-fitting. They assume that the data is I(1) and cointegrated with causality in at least one direction. This allows comparable tests of both true and false Granger noncausality hypotheses in the same model. Clarke and Mirza (2006) allow a wider variety of data-generating processes. They show that pre-testing for cointegration and then either imposing the cointegration restrictions or estimating a VAR in levels or first differences depending on the results can lead to very inflated type 1 errors in Granger causality tests. On the other hand, the Toda-Yamamoto test performed best across all data-generating processes. Analysis of our meta-dataset also shows that researchers include more lags in smaller samples and that these models have higher levels of significance *ceteris paribus*.

So, sample size can affect degrees of freedom in two different ways – smaller samples directly reduce the degrees of freedom and also encourage researchers to add lags to the regression depleting degrees of freedom further. Figure 3 illustrates this causal structure assuming that there is a genuine effect. The red channel is the statistical power relationship we want to estimate while the grey channels are the over-fitting and over-rejection pathways that we want to exclude. In our sample, it appears that the grey channel dominates and the genuine effect is weak and hence there is little effect of sample size on significance. If we include the square root of degrees of freedom in the meta-regression model while holding the degrees of freedom lost in fitting the model constant we will only measure the effect of degrees of freedom due to increases in sample size. This will eliminate the grey path in Figure 3.

$$Z_i = \alpha_0 + \alpha_1 DF_i^{0.5} + K_i + v_i \quad (8)$$

where K is the degrees of freedom lost in fitting underlying VAR. This includes the number of coefficients estimated as well as initial observations dropped because of adding lagged variables. We also test the effect of the number of lags and other variables such as time trends. When the number of lost degrees of freedom is used, α_1 models only the effect of the square root of degrees of freedom due to the direct effect of the sample size and, therefore,

---

⁶ However, most of the studies in our meta-sample use bivariate VARs

eliminates the effects of intentional and unintentional data-mining via model specification searches.

Another twist is that over-fitting has theoretically worse effects on over-rejection when the sample and degrees of freedom are small. And there is less of a problem with a large number of lags when the sample size is sufficiently large. We tried to take this into account in the empirical analysis by adding interaction terms but this had little effect.

Choice of Studies

There are a very large number of papers in the energy-output causality literature, which vary considerably in methodology, data, and econometric quality. Academic publication rewards novelty and so there are many unique studies which are hard to compare to others. As meta-analysis requires some commonality between studies, some studies must be excluded. This section describes the methods and criteria we used to select our sample of studies, which are listed in Table 1.

Two recently published surveys (Ozturk, 2010; Payne, 2010) list many relevant studies. We also searched *Scopus*, *EconLit*, and *Google Scholar* for combinations of the keywords “energy”, “electricity”, “coal”, “gas”, “oil”, “nuclear”, “GDP”, “growth”, “income”, “output”, “economy”, “causality”, “cointegration”, and “relation” to find more studies. We also include some unpublished studies in order to attempt to reduce publication bias. We collected more than five hundred papers. However, only a small subset were coded and included in the meta-analysis. We filtered papers for commensurability and econometric quality and also had to exclude papers because they did not provide all the information that was required for our meta-analysis.

Possible specifications of the energy variable are: total energy consumption, coal, electricity, natural gas, non-renewable energy, nuclear energy, oil, petrol, petroleum products as well as renewable energy sources. Possible specifications of the output variable are GDP and GNP, as well as value added from the different sectors of the economy. The variables are either related to the macroeconomic level or to single sectors of the economy such as the commercial, services, transportation, industry, residential, and agricultural sectors. Many studies test for causality between energy and output variables at different levels of aggregation, for example between national electricity use and output of the industrial sector.

These results may be spurious (Zachariadis, 2007; Gross, 2011). We included such studies but also coded a subsample of studies which use macro-level variables for both energy and output. A further sub-sample of studies within this sample is restricted to only those studies using total energy rather than individual energy carriers such as electricity or oil alone.

We include studies that use causality tests developed by Granger (1969), Sims (1972), Hsiao (1979), or Toda and Yamamoto (1995) or cointegration tests developed by Engle and Granger (1987) or Johansen (1988; 1991). For the cointegration tests we note whether the test is a test for causality in the short-run or long-run dynamics only, or a joint test. All the above approaches only include lagged values of the time series on the right hand side (RHS) of the estimated regression equations. We excluded models that include contemporaneous terms on the RHS such as the so-called instantaneous Granger causality test (e.g. Zarnikau, 1997) and the autoregressive distributed lags (ARDL) bounds test developed by Pesaran and Shin (1999) and Pesaran *et al.* (2001). The former is an inappropriate model for testing Granger causality (Granger, 1988) and the latter approach assumes the direction of Granger causality, *a priori*. We also excluded results using unique methodologies such as nonparametric approaches (e.g., Azomahou *et al.*, 2006) and threshold cointegration (e.g., Esso, 2010).⁷ For reasons of comparability all studies that found more than one cointegrating vector using the Johansen approach were excluded.⁸

The majority of studies use annual data for individual countries. We excluded studies using quarterly as well as monthly data. We also excluded studies using panel data because we constructed the database in order to also be able to test the effects of the level of economic development and other country characteristics on the direction of causality. Similarly, we exclude studies for the sub-country level, e.g., cities, regions, and provinces, for reasons of comparability. Also studies for Taiwan were excluded because information on the Taiwanese economy that is comparable to other countries is somewhat limited.

We could only include those studies, which contain all relevant information needed for the empirical tests, in particular information on the lag structure of each variable. This information is needed for calculation of the degrees of freedom. If the required information was not provided in the paper, we contacted the corresponding authors. We exclude

---

⁷ In the latter case, we coded the Toda-Yamamoto causality tests included in the paper. There are many such instances where we only partially coded a paper.

⁸ This includes Stern (2000).

potentially relevant studies if we did not receive any reply or if the answer was still incomplete.

Finally, we excluded studies if the estimation strategy is incorrect - for example different lag lengths were used for the Johansen-Juselius cointegration test and the VECM based on the estimated cointegration vectors - or if the presentation of results is unclear or statistically incorrect (e.g., negative F-statistics). This includes all Granger causality tests in levels that do not use the Toda-Yamamoto approach. A large number of early studies including Stern (1993) were thus excluded. Another example is Chang and Soruco Carballo (2011), which we excluded because only significant results were reported in the paper and one test statistic had the same value exactly in two countries. The aim of these exclusions is to reduce the effect of spurious regression or other econometric errors on the meta-analysis. We documented the reasons for exclusion for all studies. This information is available on request.

Exploratory Analysis of the Data

Description of the Data

A total of 72 studies with 1142 observations are included in the full sample. There are 574 observations of growth causes energy and 568 of energy causes growth. There are a total of 428 macro-macro only observations (425 in the energy – output direction of causality) though not all of those use aggregate energy. The number of macro-macro observations using total energy is 314 (313 in the energy – GDP direction of causality).

In all cases, we treat the effective sample size as the length of the time series despite the fact that some test statistics were produced by system estimators that use the information in all equations of the system and other test statistics are based on single equation estimation. We found it impossible to tell in many cases exactly how a model was estimated. For example, an author might say they use the Johansen procedure but in fact they just used it to estimate the cointegrating vectors. They then estimate a VECM using OLS with the error correction terms derived from the Johansen estimate.

Table 2 provides information on the distribution of the negative of the probit transformed test statistics for the full sample and a sub-sample excluding cointegration studies. The mean test statistic for energy causes output is 1.047, which is associated with a p-value of 0.148 and for output causes energy 1.153 (p=0.124). So, the average test statistics in the underlying studies

are not significant at conventional levels.⁹ Additionally, the standard deviations in Table 2 are greater than unity so that there is a more dispersed distribution than expected under the null of no causality, where we would expect the statistics to be distributed as $N(0,1)$. The four percentiles in the upper tail are also much greater than the expected values under the null of 1.28, 1.65, 1.96, and 2.32. So there is certainly a large amount of excess significance. Figures 1 and 2 compare the distribution of the test statistics to the standard normal distribution. Both have lower central frequencies with a large range of equal frequencies and a fat upper tail. This could be because:

- a. There are genuine effects in the metadata that need to be uncovered even though the majority of test statistics are not significant at traditional levels.
- b. Publication bias results in studies with more significant results being more likely to be published than those with less significant results and/or authors do not bother reporting some insignificant results and carry out specification searches to generate more significant test statistics.
- c. Spurious regression - results seem highly significant when they are not. Given our efforts to only include cointegrated studies, Toda-Yamamoto tests, or Granger causality tests in first differences in the dataset, the classic notion of spurious regression (Granger and Newbold, 1974) is probably not the cause of these results. However, as discussed above, in the typically short-time series used in this literature there are tendencies to over-fit models and for such over-fitted models to be spuriously significant.
- d. A finding of cointegration between variables implies that there is Granger causality in at least one direction (Engle and Granger, 1987). This prescreening and often inappropriate methods of testing for causality in cointegrated models means that the reported significance levels may be exaggerated (Clarke and Mirza, 2006).

To test explanation d. we also present in Table 2 statistics for samples excluding the results of cointegration tests. Though this reduces the excess significance, there is still a lot of excess significance that needs to be explained. We will test explanations a., b., and c. in the

---

⁹ On the other hand, these means are significantly greater than the zero value expected under the null hypothesis. The t-statistics for the difference of the means from zero are 15.8 and -16.8 respectively.

remainder of the paper while controlling for the effect of cointegration pre-screening.

Figures 4 and 5 illustrate the distribution of the individual test statistics plotted against the square root of degrees of freedom – a version of the Galbraith plot (Stanley, 2005a). The dotted line is for a test statistic value of 1.65. The outliers to the right in each figure are from Vaona (2012). There does not seem to be a strong relationship between degrees of freedom and the size of the test statistics. The figures do show that the test statistics are fairly evenly distributed around the mean and that there are a very large number of test statistics greater than the 5% significance level.

Correlation Analysis

The correlations of most interest are between the key dependent and explanatory variable in the MST regression – the test statistics and the square root of degrees of freedom and the other variables in the data set as well as between sample size and all the other variables. These correlations are presented in Table 3.

The correlations are mostly pretty similar for the causality tests in each direction. The correlations between the test statistics and the square root of degrees of freedom are negative but weak (-0.055 and -0.013). This is the opposite of the expected relationship if there were real effects in the studies. There are very weak positive relations between the test statistics and sample size. But the number of coefficients in the regression (KEG and KGE)¹⁰ is positively associated with the test statistics (significant at 0.1% level for E-G and 5% for G-E). The test statistics are significantly higher in studies that find cointegration as we found above. This makes sense, as this is a pre-screening for Granger causality in at least one direction. Studies that include capital (which includes gross fixed capital formation as well as the capital stock) or employment are more significant. Later sample start dates are positively but weakly associated with the test statistic as are later sample end points and the publication year. So it seems from the latter that the relationship between energy and growth may have strengthened over time though of course this does not control for changes in methodology and in the sample of countries.

---

¹⁰ This, of course, is the total number of variables in the regression. But the latter term could be confusing because it might refer to the number of different time series in the VAR, which we designate by the variable “VARIABLES”. KEG and KGE count each lag of each variable as well as the constant and time trend if present and are computed as the difference between the sample size and the degrees of freedom.

As we would expect, degrees of freedom is negatively correlated with the start date but is much more weakly (but still highly significantly) positively associated with the end year of the sample. Degrees of freedom are also strongly negatively correlated with the number of coefficients. At first glance, this might appear to make sense – increasing the number of variables reduces the degrees of freedom. But that is only true holding the sample size constant! Usually, as the sample size grows, both the number of variables and the degrees of freedom will increase if researchers add extra variables at a slower rate than they increase the sample size. But in fact sample size is also somewhat negatively correlated with the total number of regression coefficients in each regression. These phenomena can be explained by hypothesis c. As explained above, there is a tendency to over-fit models in small samples and for these models to have inflated type 1 errors. There is also a negative correlation between the sample size and the number of lags and the presence of a time trend. Sample size is, however, positively associated with the number of controls – variables other than energy and output - as well as with the specific controls of capital and energy prices.

Basic Meta-Regression Analysis

Table 4 presents the results for the full sample for the basic meta-regression models. The first three columns for both directions of causation are the simple meta-regression models and the second three columns control for cointegration studies. Results are remarkably similar in both directions except the energy to growth direction is generally slightly more significant (higher R-squared). All the slope coefficients are negative, though only one of the probit transform models has a statistically significant slope. Though we might expect a negative slope for the t-statistics as explained above, it is unexpected to find a negative slope for the other two forms of the test statistics. The negative slope is reduced by using normal variates instead of t-statistics and even more by not taking absolute values and logarithms. The cointegration dummy has a large positive and statistically significant coefficient in all models. For energy causes growth the combined intercept is 2.22 for the cointegration models and for growth causes energy 1.876 which is associated with the 3% significance level. Screening for cointegration should result in significant Granger causality in at least one direction, on average it is found in both directions.

Exactly as we would expect if there were no genuine effect in the data, we cannot reject the null of homoskedasticity for any of the models at the 5% level. However, the residuals from the logarithmic models are highly non-normal. The residuals from the probit transform model are still non-normal in the growth causes energy direction but the test statistics are much smaller than

for the logarithmic models. Therefore, we only use the probit transform model in the remainder of the paper. The intercept term of the probit transform model is highly significant, suggesting publication or misspecification bias.

We also test for the effect of the observations from Vaona (2012), which constitute an outlier in terms of degrees of freedom. The coefficients change very little when these observations are removed, so they are not the cause of the negative slope.

Explaining the Negative Slope

Alternative Hypotheses

Card and Krueger (1995) also found a negative relationship between the variables in an MST regression and suggested that this could be due to either publication bias or changes in the minimum wage relationship over time. They preferred the publication bias hypothesis based on further tests. MST is, however, very powerful in the face of a uniform publication bias (Stanley, 2008). So there would have to be more publication bias in studies with fewer degrees of freedom than is necessary to simply obtain significant results in order for a negative relationship to result. We think the most likely explanation for the negative slope that we find in our study is the over-fitting over-rejection hypothesis. But we also examine other three alternative hypotheses:

1. The significance of the relationship between energy and growth may have declined over time and studies with fewer degrees of freedom represent studies from an earlier period, whereas studies with more degrees of freedom represent datasets that include more recent data. The relatively low correlation between end date and sample size and the positive relation between both start and end dates and the size of the test statistics (Table 3) suggests that this is not the case – more recent data is likely to have a higher test statistic.

We can also test this with regression II in Table 5. Holding the sample size constant and increasing the end date, effectively moves a time window of fixed length through the data. The results show that increasing the end date has a positive (though only in one case significant) effect on the reported test statistics. So this rejects the hypothesis that the test statistics are smaller in more recent samples. Controlling for the end point

and sample size also results in a more negative and significant effect of the degrees of freedom. As the sample size is held constant, this now measures the effect of removing parameters from the model, through removing control variables, lags, and deterministic components. The more of these that are removed the higher the degrees of freedom and the less significant the test statistic. Dropping the end point variable (Regression I) has relatively little effect on the latter phenomenon showing the difference between the effects of sample size and degrees of freedom generally.

2. There may be more changes in the economy over longer periods and, therefore, the effects of energy on growth or *vice versa* may be obscured as the size of the sample gets larger. This is a generalization of hypothesis 1. We see in Regression IV in Table 5 that when end year and number of parameters are held constant, sample size has no effect on the dependent variable. As this is now the pure effect of the length of sample with the time period and number of parameters controlled for, this hypothesis cannot explain the negative slope of degrees of freedom. Also, from Table 3 we see that sample size has a very weak positive simple correlation with the test statistics when we do not control for other variables. Therefore, there is no strong evidence for this hypothesis.
3. In the presence of publication bias, studies with fewer degrees of freedom need to select for large effect sizes in order to obtain significant results. However, even if results from low degrees of freedom studies are more significant than they should be, larger studies should still get more significant results if there is a real effect. And if there is no genuine effect the publication bias should be uniform across degrees of freedom. However, authors with smaller samples could be more prone to trying to get significant results than authors with larger samples. If there were no genuine effect the slope of degrees of freedom would be negative. Though this is possible, it is not testable.

Exploring Misspecification Bias

As we saw in the correlation analysis (Table 3), sample size is somewhat negatively correlated with the number of degrees of freedom lost in model fitting (KEG and KGE). Usually, we would expect that as the sample gets larger, researchers are able to add more variables to their regression. But here we see the reverse. There is also a negative correlation

between the sample size and the number of lags and the presence of a time trend, though a positive correlation between sample size and the numbers of control variables. The number of lags is very negatively correlated with the degrees of freedom (Table 3). So researchers with small samples tend to add a lot of lags, which greatly deplete the degrees of freedom. The number of lags of energy in the energy causes growth tests are significantly positively correlated with the test statistic for these tests. The number of lags of output in the growth causes energy equation are positively correlated with the Z-score for those tests though this correlation is not significant at the 10% level. We also found in the previous section (Table 5 column I) that when controlling for sample size, degrees of freedom has a negative effect on the test statistics and a more negative effect than when we do not control for sample size (Table 4). This is good evidence in favor of the over-fitting over-rejection hypothesis.

We explore further this potential effect in our data using the regressions reported in Columns III to VI in Table 5. Degrees of freedom is the difference between the original sample size and the number of regression coefficients estimated and initial observations dropped. So, in this section our base line model (III) uses degrees of freedom in levels rather than the square root in order to be able to decompose degrees of freedom into these two components and see their effect on the test statistics. The basic model (Table 5, Column III) shows similar results to the equivalent regression in Table 4. The residual properties also change little.

We then split DF into SAMPLE and KEG or KGE - to show that the number of coefficients and dropped initial observations is the main driver of the negative coefficient on degrees of freedom (Column IV). If we increase SAMPLE with KEG or KGE held constant we will see the effect of DF on the Z-scores due to an increase in sample size. Sample size does not have a significant effect on the Z-scores, *ceteris paribus*, while the number of coefficients has a significant positive effect. This result strongly supports our hypothesis. In columns IV to VI the variables for coefficient, lags, and control numbers are demeaned, so that the intercept term is for a study with average numbers of these. It is hard to specifically identify publication bias rather than misspecification bias, but the intercept is much reduced for these three models and insignificant, suggesting that much of what appears to be publication bias in the simple meta-regression model is in fact due to misspecification bias.

In Column V, we add the various types of variables that can be included in the underlying VAR models.¹¹ We also add a dummy for the Hsiao procedure because this approach results in different numbers of lags for the different variables. Now KEG or KGE has an insignificant effect showing that the additional variables explain most of the effect. Dropping the KEG or KGE (Column VI) produces similar results and appears to reduce multicollinearity. Therefore, we focus on these results. The number of lags of energy is significant in the energy causes growth tests but lags of output are not. Neither lags variable is significant in the growth causes energy tests at conventional significance levels. The number of controls is significant in the energy causes growth equation ($p=0.057$). Time trends have a large and significant effect in the energy causes growth tests. Of course, adding a time trend is not necessarily a misspecification but it clearly affects the results. Using the Hsiao procedure increases significance for both energy causes growth and growth causes energy. This makes sense as it selects the number of lags of energy to deliberately get the most significant fit. We also tried dropping one of the lags variables from each equation but this made little difference to the results.

From this it is clear that, in the full sample, the portion of degrees of freedom that is not affected by model fitting has no effect on significance and, therefore, there are no observable real effects in this literature as a whole. It is still possible that some studies that find no significant effect overall, then split their datasets up and if they find a significant result, report that, contaminating this variable too with publication bias when in fact there are real effects. We test this hypothesis by running regression IV using only those studies that report a single sample size.¹² These regressions (for either energy causes growth or *vice versa*) do not produce a significant positive coefficient for sample size. Therefore, such contamination does not appear to be a problem.

It is interesting that lags of energy and time trends have significant effects on the test statistics, whereas the number of control variables does not and that the number of controls is

---

¹¹ Note that the total number of coefficients in the regression is equal to 2 plus the number of controls times the average number of lags of the variables plus the number of deterministic components. Therefore, the total number of coefficients can be retained in the regression to test the effect of this non-linearity.

¹² As not all results from studies included in our sample were coded we rechecked the original papers to make sure that in each case the authors used data from a single sample period only. This sample also excludes studies that have multiple sample sizes due to the differing availability of data for different countries.

positively associated with sample size. This suggests that control variables are not added to regressions to obtain significant results whereas lags and time trends are. In fact, for the subsample that uses control variables there is no correlation (less than 0.01) between the number of lags and the sample size. This fits the finding of Gonzalo and Pitarakis (2003) that over-fitting is less likely in higher dimensional VARs.

All this evidence strongly supports the over-fitting over-rejection hypothesis. Table 6 presents estimates of versions of the meta-Granger causality model (8). Degrees of freedom has a positive coefficient in five of the six regressions in the table but is not significant at conventional levels. Therefore, we conclude that there is no observable genuine effect in the meta-sample as a whole. The effect is larger in the models that control for total coefficients (A) rather than just lags (B). Model C adds some of the other variables from Table 5, improving performance further. As the intercept is insignificantly different from zero, over-fitting, cointegration pre-screening, and inclusion of time trends can largely explain the excess significance.

Effects of Methodology on Finding a Genuine Effect

Though we cannot find a significant real Granger causality effect in the sample as a whole, perhaps some methodological approaches do make a difference and uncover real causality effects. Ozturk (2010) and Stern (2011) both argue that some methods are more likely to uncover a robust effect. In this section we test for whether there are any methodologies where a genuine effect can be found. These include both econometric methods and the inclusion of various control variables. This is tested by adding a dummy variable and an interaction term between the dummy and the degrees of freedom variable to a basic version of the model:

$$Z_i = \alpha_0 + \alpha_1 DF_i^{0.5} + \alpha_2 K_i + \beta_0 d_i + \beta_1 d_i DF_i^{0.5} + v_i \quad (9)$$

where d is the dummy variable that equals 1 if the methodology was employed. We drop the cointegration dummy because that would confuse interpretation of the results for the different methodologies. Table 7 reports coefficient values and t-tests for $\alpha_1 + \beta_1$ and $\alpha_0 + \beta_2$ only. The former is a test for a genuine effect when the methodology in question is used and the latter is a test of whether there is excess significance when the method is used. For some

methodologies of interest we have insufficient data to test these hypotheses. For example, Oh and Lee (2004) is the only paper in our sample to use quality-adjusted energy.

The majority of methodologies that we tested do not have significant genuine effects. Where we do find genuine effects these indicate that GDP causes energy. First we test the various techniques. There do appear to be genuine effects for cointegrated results. The result in the growth causes energy direction is significant at the 5% level in a one-tailed test. However, tests on the short-run coefficients from cointegrated VARs are not significant and only long-run or joint long and short-run tests are significant and then only in the growth causes energy direction. Results are particularly significant for the Engle-Granger technique. However, there is excess significance for this technique in the energy causes growth direction.

Traditional Granger causality tests – in first differences – do not show a genuine effect and have excess significance for growth causes energy. The Hsiao and Toda-Yamamoto tests have a large amount of excess significance.

Among the variables, only those models with energy prices have a significant genuine effect for GDP causes energy. This model defines a demand function where energy use is determined by prices and income rather than the production function relationship that would be determined if energy causes output. Models that include capital have a large amount of excess significance.

The macro-macro subsample may have a genuine effect from output to energy (one-tailed test $p\text{-value} = 0.054$). This possibly extends the validity of Gross' (2011) findings. Further restricting the sample to total energy only, reduces the significance of this effect. We repeated all the tests in Table 7 using only the macro-macro subset of data. Results are very similar in this subset though price was less statistically significant though its coefficient was only slightly smaller than in Table 7. We also tested for effects in a sample excluding the cointegration studies. These results were also similar to those in Table 7.

We also estimated models that included the effect of multiple variables. For example, we included effects for cointegration, Toda-Yamamoto, and the Hsiao procedure, treating simple Granger causality as the default. We also estimated this model splitting the cointegration category into Engle-Granger and Johansen methods and short-run, long-run, and joint tests. None of these tests of genuine effects or excess significance was different to those in Table 7 in terms of sign or significance level models. We also estimated a model with effects for

TIME, PRICE, and CAPITAL with similar results. Including the CONTROLS variable in the regression as well though removed the significance of PRICE though when we included CONTROLS but not CAPITAL, PRICE has a significant effect.

Finally, we tested joint hypothesis of whether there are genuine effects when using particular control variables with specific methods:

$$Z_i = \alpha_0 + \alpha_1 DF_i^{0.5} + \alpha_2 K_i + \beta_0 m_i + \beta_1 m_i DF_i^{0.5} + \sum_j (\gamma_{0j} c_{ji} + \gamma_{1j} c_{ji} DF_i^{0.5} + \gamma_{2j} m_i c_{ji} DF_i^{0.5}) + v_i \quad (10)$$

where m is the dummy variable for a method and the c_j are dummies for the various possible control variables in the underlying studies. The interaction term between the two dummies and the square root of degrees of freedom tests if there is a difference in genuine effect using this method when the control variable in question is present in the study. One could also add an interaction between the two dummies alone, but we found these effects to be insignificant and dropped them. This model is estimated separately for each method.

The results are reported in Table 8 in terms of t-statistics for linear combinations of regression coefficients that measure the stated treatments. The first row tests $\alpha_1 + \beta_1 = 0$ - which is a test of a genuine effect when the named method is used but no control variables are included. The second row, and similar rows for other control variables, tests $\alpha_1 + \beta_1 + \gamma_1 + \gamma_2 = 0$, which tests whether there is genuine effect when this method and control is used, setting all other controls to zero. Other controls are all control variables apart from capital, employment, and price. Carbon dioxide emissions are the most important of these.

When no control variables or a time trend are included cointegration in general and the Johansen procedure and joint short- and long-run causality tests appear to have a genuine effect in the energy causes GDP direction. Long-run and joint tests and the Engle-Granger procedure have genuine effects in the growth causes energy direction. None of the significant effects in the energy causes growth direction hold up when either capital or prices are added to the models. This suggests that they are due to omitted variables bias. Adding time trends or employment however, increases the significance of the “genuine” effects. We also jointly tested whether there was a genuine effect when capital, employment, and a time trend are included as in Stern (2000). These results are in the last line of the energy causes growth panel of Table 8. We do find a genuine effect for this model when using the Johansen

procedure. But the results are less significant than when only employment is present. A possible explanation is that the elasticity of substitution between capital and energy is small and, therefore, the movements of energy, while holding energy constant are small too and have insignificant effects on output (Rotemberg and Woodford, 1996). Similarly, the reason why we find a stronger effect from growth to energy rather than *vice versa* is because the share of energy in output is small while the role of income in energy demand is larger.

In the growth causes energy direction the significant effects are no longer present when capital is added to the model either. But adding prices strengthens the effect. For the Johansen procedure there is no significant effect unless prices are added. This is the energy demand function model, which is supported by economic theory without necessarily including a capital variable.

Discussion and Recommendations

A very large literature has developed that uses time series analysis to test whether energy causes economic output or vice versa with little in the way of conclusive results or guidance on how to model relationships between energy and economic output. This paper provides the first meta-analysis of this literature that tests whether these results are largely spurious outcomes of misspecification and publication selection biases or whether genuine statistically significant effects exist in this literature.

We find that models that include energy prices as a control variable find a genuine effect from output to energy use in the long-run. A genuine causal effect also seems apparent from energy to output when employment is controlled for and the Johansen procedure is used. This effect is more ambiguous because is only present when cointegration test screening is used and cointegration found whereas we find an effect from growth to energy when price is present across all Granger causality test methods. The finding of a robust energy demand function relationship is in line with the conclusions of Stern's (2011) literature review. Stern (2011) also argued that VAR models of quality-adjusted energy, capital, and output were likely to find that energy caused output. We could not test the effects of using quality-adjusted energy in this study due to only having one such study in our sample. The finding that when we control for employment energy causes output is in line with this conclusion but only partly as controlling for capital reduces the significance of the effect.

We did not find any genuine effects in results from Toda-Yamamoto causality tests, which should be more appropriate than tests on cointegrated VARs. The cointegrated VAR results have already been pre-screened for cointegration. Cointegration implies Granger causality in at least one direction and, therefore, it is not surprising that we find it in this subset of the literature. However, the significance of the test statistics increases with the degrees of freedom and so this effect does appear to be real.

We also found that there may be causality from output to energy more generally in the subset of the literature using only macro-level data.¹³ This extends Gross' (2011) finding that only when variables at the same level of aggregation are included in a time series model can Granger causality be found. However the significance level for this test was 5.5% in a one-tailed test and so is not extremely reliable. This finding is worthy of further future exploration.

Therefore, the only really solid finding is that when energy prices are included in VAR models it is found that output causes energy use though there are signs of significant effects in subsamples of the literature. Future research should include more studies using quality adjusted energy, more studies with very long time series – we only have one study with a time series with more than one hundred observations - and more investigation of subsets of the data with consistently defined variables. Also, studies using panel data should be investigated as these were deliberately excluded from the current study.

The meta-Granger causality tests used in this paper could also be applied in other research literatures where Granger causality testing has been common. We have some general recommendations for such future studies. We recommend to convert all Granger causality test statistics to normal variates using the negative of the probit transformation and to include control variables in the degrees of freedom lost in fitting the model to counteract the tendency to over-fit VAR models in small samples, which leads to inflated type 1 errors. We find that such models find possible genuine causality effects in some subsamples of our meta-data. The coefficients of the power trace are positive but not significant in the full sample. We also show that traditional logarithmic meta-significance (MST) models have very non-normal residuals. There is no good reason to use these models as we show that the FAT-PET model can be motivated by both statistical power and publication bias arguments. MST is simply a logarithmic version of FAT-PET. The slope-coefficient in the weight least squares version of

---

¹³ Rather than data for individual industries or mixed aggregate and sub-industry level data.

FAT-PET measures the genuine effect by exploiting the power trace while the constant is a test of publication bias.

We confirmed the finding in the econometric literature that there is a tendency to over-fit the number of lags of the time series in small samples and that these over-fitted models tend to over-reject the null hypothesis when it is true. All models without the control variables have a negative coefficient on the power trace function, which is most pronounced if we convert test statistics to t-statistics and then take logs of the absolute values. Even where the original test statistics are t distributed it is better to convert them to normal variates.

References

- Abosedra, S., Baghestani, H., 1991. New Evidence on the Causal Relationship between United States Energy Consumption and Gross National Product. *Journal of Energy and Development* 14 (2), 285-292.
- Acaravici, A., 2010. Structural Breaks, Electricity Consumption and Economic Growth: Evidence from Turkey. *Romanian Journal of Economic Forecasting* 2, 140-154.
- Adom, P.K., 2011. Electricity Consumption-Economic Growth Nexus: The Ghanaian Case. *International Journal of Energy Economics and Policy* 1 (1), 18-31.
- Akinlo, A.E., 2008. Energy Consumption and Economic Growth: Evidence from 11 Sub-Saharan African Countries. *Energy Economics* 30 (5), 2391-2400.
- Akinlo, A.E., 2009. Electricity Consumption and Economic Growth in Nigeria: Evidence from Cointegration and Co-Feature Analysis. *Journal of Policy Modeling* 31 (5), 681-693.
- Alam, M., Begum, I., Buysse, J., Rahman, S., Van Huylenbroeck, G., 2011. Dynamic Modeling of Causal Relationship between Energy Consumption, CO2 Emissions and Economic Growth in India. *Renewable and Sustainable Energy Reviews* 15 (6), 3243-3251.
- Altinay, G., Karagol, E., 2005. Electricity Consumption and Economic Growth: Evidence from Turkey. *Energy Economics* 27 (6), 849-856.
- Ang, J. B., 2008a. A Survey of Recent Developments in the Literature of Finance and Growth. *Journal of Economic Surveys* 22(3), 536-576.
- Ang, J.B., 2008b. Economic Development, Pollutant Emissions and Energy Consumption in Malaysia. *Journal of Policy Modeling* 30 (2), 271-278.
- Azomahou, T., Laisney, F., Nguyen Van, P., 2006. Economic Development and CO2 Emissions: A Nonparametric Panel Approach. *Journal of Public Economics* 90 (6-7), 1347-1363.
- Belloumi, M., 2009. Energy Consumption and GDP in Tunisia: Cointegration and Causality Analysis. *Energy Policy* 37 (7), 2745-2753.
- Boehm, D., 2008. Electricity Consumption and Economic Growth in the European Union: A

- Causality Study Using Panel Unit Root and Cointegration Analysis. *EEM 2008, 5th International Conference on European Electricity Market*, IEEE Publications pp. 1-6.
- Bowden, N., Payne, J., 2009. The Causal Relationship between US Energy Consumption and Real Output: A Disaggregated Analysis. *Journal of Policy Modeling* 31 (2), 180-188.
- Bressler, S. L. and Seth, A. K., 2011. Wiener-Granger Causality: A Well Established Methodology. *NeuroImage* 58(2), 323-329.
- Card, D., Krueger, A. B., 1995. Time-Series Minimum-Wage Studies: A Meta-Analysis. *American Economic Review* 85 (2), 238-243.
- Chang, C., Soruco Carballo, C., 2011. Energy Conservation and Sustainable Economic Growth: The Case of Latin America and the Caribbean. *Energy Policy* 39 (7), 4215-4221.
- Chebbi, H., 2009. Investigating Linkages between Economic Growth, Energy Consumption and Pollutant Emissions in Tunisia. *International Association of Agricultural Economists*.
- Chen, P.-Y., S.-T. Chen, and Chen, C.-C., 2012. Energy Consumption and Economic Growth—New Evidence from Meta Analysis. *Energy Policy* 44, 245-255.
- Chiou-Wei, S., Chen, C., Zhu, Z., 2008. Economic Growth and Energy Consumption Revisited— Evidence from Linear and Nonlinear Granger Causality. *Energy Economics* 30 (6), 3063-3076.
- Chontanawat, J., Hunt, L., Pierse, R., 2008. Does Energy Consumption Cause Economic Growth? Evidence from a Systematic Study of Over 100 Countries. *Journal of Policy Modeling* 30 (2), 209-220.
- Ciarreta, A., Otaduy, J., Zarraga, A., 2009. Causal Relationship between Electricity Consumption and GDP in Portugal: A Multivariate Approach. *The Empirical Economics Letters* 8 (7), 693-701.
- Clarke, J.A., Mirza, S., 2006. A Comparison of Some Common Methods for Detecting Granger Noncausality. *Journal of Statistical Computation and Simulation* 76, 207-231.
- Engle, R.E., Granger, C.W.J., 1987. Cointegration and Error-Correction: Representation, Estimation, and Testing. *Econometrica* 55, 251-276.
- Erol, U., Yu, E., 1987. Time Series Analysis of the Causal Relationships between US Energy and Employment. *Resources and Energy* 9 (1), 75-89.
- Esso, L. J., 2010. Threshold Cointegration and Causality Relationship between Energy Use and Growth in Seven African Countries. *Energy Economics* 32 (6), 1383-1391.
- Fallahi, F., 2011. Causal Relationship between Energy Consumption (EC) and GDP: A Markov-Switching (MS) Causality. *Energy* 36 (7), 4165-4170.
- Ghosh, S., 2002. Electricity Consumption and Economic Growth in India. *Energy Policy* 30 (2), 125-129.
- Glasure, Y.U., 2002. Energy and National Income in Korea: Further Evidence on the Role of Omitted Variables. *Energy Economics* 24 (4), 355-365.
- Glasure, Y.U., Lee, A.-R., 1997. Cointegration, Error-Correction, and the Relationship between GDP and Energy: The Case of South Korea and Singapore. *Resource and Energy Economics* 20 (1), 17-25.

- Golam Ahamad, M., Nazrul Islam, A., 2011. Electricity Consumption and Economic Growth Nexus in Bangladesh: Revisited Evidences. *Energy Policy* 39 (10), 6145-6150.
- Gonzalo, J., Pitarakis, J.-Y., 2003. Lag Length Estimation in Large Dimensional Systems. *Journal of Time Series Analysis* 23(4), 401-423.
- Granger, C. W. J., 1969. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37 (3), 424-438.
- Granger, C. W. J., 1988. Some Recent Developments in a Concept of Causality. *Journal of Econometrics* 39, 199-211.
- Granger, C.W.J., Newbold, P., 1974. Spurious Regressions in Econometrics. *Journal of Econometrics* 2, 111-120.
- Gross, C. (2011). Explaining the (Non-) Causality between Energy and Economic growth in the U.S.—A Multivariate Sectoral Analysis. *Energy Economics* 34(2), 489-499.
- Hacker, R.S. and Hatemi-J, A. 2008. Optimal Lag-Length Choice in Stable and Unstable VAR Models under Situations of Homoscedasticity and ARCH. *Journal of Applied Statistics* 35 (6), 601-615.
- Hamilton, J.D., 2009. Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity* 2009 (1), 215-261.
- Hondroyannis, G., Lolos, S., Papapetrou, E., 2002. Energy Consumption and Economic Growth: Assessing the Evidence from Greece. *Energy Economics* 24 (4), 319-336.
- Hsiao, C., 1979. Autoregressive Modeling of Canadian Money and Income Data. *Journal of the American Statistical Association* 74, 553-560.
- Ioannidis, J. P. A., 2005. Why Most Published Research Findings Are False. *PLoS Medicine* 2(8), e124.
- Jamil, F., Ahmad, E., 2010. The Relationship between Electricity Consumption, Electricity Prices and GDP in Pakistan. *Energy Policy* 38 (10), 6016-6025.
- Jamil, F., Ahmad, E., 2011. Income and Price Elasticities of Electricity Demand: Aggregate and Sector-Wise Analyses. *Energy Policy* 39 (9), 5519-5527.
- Jobert, T., Karanfil, F., 2007. Sectoral Energy Consumption by Source and Economic Growth in Turkey. *Energy Policy* 35 (11), 5447-5456.
- Johansen, S., 1988. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12 (2-3), 231-254.
- Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 59 (6), 1551-1580.
- Jumbe, C., 2004. Cointegration and Causality between Electricity Consumption and GDP: Empirical Evidence from Malawi. *Energy Economics* 26 (1), 61-68.
- Kaplan, M., Ozturk, I., Kalyoncu, H., 2011. Energy Consumption and Economic Growth in Turkey: Cointegration and Causality Analysis. *Journal for Economic Forecasting* (2), 31-41.
- Karanfil, F., 2008. Energy Consumption and Economic Growth Revisited: Does the Size of

- Unrecorded Economy Matter? *Energy Policy* 36 (8), 3029-3035.
- Kaufmann, R. K. and Stern, D. I., 1997. Evidence for Human Influence on Climate from Hemispheric Temperature Relations. *Nature* 388, 39-44.
- Lee, C., 2006. The Causality Relationship between Energy Consumption and GDP in G-11 Countries Revisited. *Energy Policy* 34 (9), 1086-1093.
- Lee, T.-H., Yang, W., 2012. Money-Income Granger-Causality in Quantiles. *Advances in Econometrics* 30, 385-409.
- Lorde, T., Waithe, K., Francis, B., 2010. The Importance of Electrical Energy for Economic Growth in Barbados. *Energy Economics* 32 (6), 1411-1420.
- Lotfalipour, M., Falahi, M., Ashena, M., 2010. Economic Growth, CO2 Emissions, and Fossil Fuels Consumption in Iran. *Energy* 35 (12), 5115-5120.
- Masih, A. M. M., Masih, R., 1996. Energy Consumption, Real Income and Temporal Causality: Results from a Multi-Country Study Based on Cointegration and Error-Correction Modelling Techniques. *Energy Economics* 18 (3), 165-183.
- Masih, A. M. M., Masih, R., 1998. A Multivariate Cointegrated Modelling Approach in Testing Temporal Causality between Energy Consumption, Real Income and Prices with an Application to Two Asian LDCs. *Applied Economics* 30 (10), 1287-1298.
- Mehrara, M., 2007. Energy Consumption and Economic Growth: The Case of Oil Exporting Countries. *Energy Policy* 35 (5), 2939-2945.
- Menyah, K., Wolde-Rufael, Y., 2010a. CO2 Emissions, Nuclear Energy, Renewable Energy and Economic Growth in the US. *Energy Policy* 38 (6), 2911-2915.
- Menyah, K., Wolde-Rufael, Y., 2010b. Energy Consumption, Pollutant Emissions and Economic Growth in South Africa. *Energy Economics* 32 (6), 1374-1382.
- Mozumder, P., Marathe, A., 2007. Causality Relationship between Electricity Consumption and GDP in Bangladesh. *Energy Policy* 35 (1), 395-402.
- Nickelsburg, G., 1985. Small-Sample Properties of Dimensionality Statistics for Fitting VAR Models to Aggregate Economic Data. *Journal of Econometrics* 28, 183-192.
- Oh, W., Lee, K., 2004. Causal Relationship between Energy Consumption and GDP Revisited: The Case of Korea 1970-1999. *Energy Economics* 26 (1), 51-59.
- Ozturk, I., 2010. A literature survey on energy-growth nexus. *Energy Policy* 38 (1), 340-349.
- Pao, H., Tsai, C., 2011. Modeling and Forecasting the CO2 Emissions, Energy Consumption, and Economic Growth in Brazil. *Energy* 36 (5), 2450-2458.
- Paul, B., Uddin, G., 2010. Energy and Output Dynamics in Bangladesh. *Energy Economics* 33 (3), 480-487.
- Paul, S., Bhattacharya, R. N., 2004. Causality between Energy Consumption and Economic Growth in India: A Note on Conflicting Results. *Energy Economics* 26 (6), 977-983.
- Payne, J.E., 2009. On the Dynamics of Energy Consumption and Output in the US. *Applied Energy* 86 (4), 575-577.
- Payne, J.E., 2010. Survey of the International Evidence on the Causal Relationship between

- Energy Consumption and Growth. *Journal of Economic Studies* 37 (1), 53-95.
- Payne, J.E., 2010. Survey of the international evidence on the causal relationship between energy consumption and growth. *Journal of Economic Studies* 37 (1), 53-95.
- Pesaran, M. H., Shin, Y., 1999. An Autoregressive Distributed Lag Modelling Approach to Cointegrated Analysis. In: Strom, S. (Ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Cambridge University Press, Cambridge, MA.
- Pesaran, M. H., Shin, Y., Smith, R. J., 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* 16 (3), 289-326.
- Pradhan, R., 2010. Energy Consumption-Growth Nexus in SAARC Countries: Using Cointegration and Error Correction Model. *Modern Applied Science* 4 (4), 74-90.
- Rafiq, S., Salim, R., 2011. The Linkage Between Energy Consumption and Income in Six Emerging Economies of Asia: An Empirical Analysis. *International Journal of Emerging Markets* 6 (1), 50-73.
- Rotemberg, J. J. and M. Woodford (1996) Imperfect Competition and the Effects of Energy Price Increases on Economic Activity. *Journal of Money, Credit, and Banking* 28 (4), 549-577.
- Sa'ad, S., 2010. Energy Consumption and Economic Growth: Causality Relationship for Nigeria. *OPEC Energy Review* 34 (1), 15-24.
- Salim, R., Rafiq, S., Hassan, A., 2008. Causality and Dynamics of Energy Consumption and Output: Evidence from Non-OECD Asian Countries. *Journal of Economic Development* 33 (2), 1-26.
- Sari, R., Soytas, U., 2009. Are Global Warming and Economic Growth Compatible? Evidence from Five OPEC Countries. *Applied Energy* 86 (10), 1887-1893.
- Shiu, A., Lam, P.-L., 2004. Electricity Consumption and Economic Growth in China. *Energy Policy* 32 (1), 47-54.
- Sims, C., 1972. Money, Income, and Causality. *American Economic Review* 62 (4), 540—552.
- Soytas, U., Sari, R., 2009. Energy Consumption, Economic Growth, and Carbon Emissions: Challenges Faced by an EU Candidate Member. *Ecological Economics* 68 (6), 1667-1675.
- Soytas, U., Sari, R., Ewing, B., 2007. Energy consumption, income, and carbon emissions in the United States. *Ecological Economics* 62 (3-4), 482-489.
- Stanley, T.D., 2005a. Beyond Publication Bias. *Journal of Economic Surveys* 19 (3), 309-345.
- Stanley, T. D., 2005b. Integrating the Empirical Tests of the Natural Rate Hypothesis: A Meta-Regression Analysis. *Kyklos* 58(4), 611-634.
- 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), 103-127.
- Stanley, T.D., Doucouliagos, H., 2011. Meta-Regression Approximations to Reduce

- Publication Selection Bias. *Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Economics Series 2011_4*.
- Stanley, T.D., Doucouliagos, H., 2012. *Meta-Regression Analysis in Economics and Business*. Routledge.
- Stern, D. I., 1993. Energy Use and Economic Growth in the USA: a Multivariate Approach. *Energy Economics* 15, 137-150.
- Stern, D. I., 2000. A Multivariate Cointegration Analysis of the Role of Energy in the U.S. Macroeconomy. *Energy Economics* 22, 267-283.
- Stern, D. I., 2011. The Role of Energy in Economic Growth. *Annals of the New York Academy of Sciences* 1219, 26-51.
- Toda, H.Y., Yamamoto, T., 1995. Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics* 66 (1-2), 225-250.
- Türkekul, B., Unakitan, G., 2011. A Co-Integration Analysis of the Price and Income Elasticities of Energy Demand in Turkish Agriculture. *Energy Policy* 39 (5), 2416-2423.
- Vaona, A., 2012. Granger Non-Causality between (Non)renewable Energy Consumption and Output in Italy since 1861: The (Ir)relevance of Structural Breaks. *Energy Policy* 45, 226-236.
- Vecchione, G., 2011. Economic Growth, Electricity Consumption and Foreign Dependence in Italy Between 1963—2007. *Energy Sources, Part B: Economics, Planning, and Policy* 6 (3), 304-313.
- Vlahinic-Dizdarevic, N., Zikovic, S., 2010. The Role of Energy in Economic Growth: The Case of Croatia. *Journal of Economics and Business* 28 (1), 35-60.
- Wolde-Rufael, Y., 2009. Energy Consumption and Economic Growth: The Experience of African Countries Revisited. *Energy Economics* 31 (2), 217-224.
- Wolde-Rufael, Y., 2010a. Bounds Test Approach to Cointegration and Causality between Nuclear Energy Consumption and Economic Growth in India. *Energy Policy* 38 (1), 52-58.
- Wolde-Rufael, Y., 2010b. Coal Consumption and Economic Growth Revisited. *Applied Energy* 87 (1), 160-167.
- Wolde-Rufael, Y., Menyah, K., 2010. Nuclear Energy Consumption and Economic Growth in Nine Developed Countries. *Energy Economics* 32 (3), 550-556.
- Yoo, S.-H., 2006. The Causal Relationship between Electricity Consumption and Economic Growth in the ASEAN Countries. *Energy Policy* 34 (18), 3573-3582.
- Yoo, S., Ku, S., 2009. Causal Relationship between Nuclear Energy Consumption and Economic Growth: A Multi-Country Analysis. *Energy Policy* 37 (5), 1905-1913.
- Yoo, S., Kwak, S., 2010. Electricity Consumption and Economic Growth in Seven South American Countries. *Energy Policy* 38 (1), 181-188.
- Yu, E.S.H., Hwang, B.-K., 1984. The Relationship between Energy and GNP: Further Results. *Energy Economics* 6 (3), 186-190.
- Yu, E., Choi, J.-Y., 1985. The Causal Relationship between Energy and GNP: An International Comparison. *The Journal of Energy and Development* 10 (2), 249-272.

- Yuan, J., Zhao, C., Yu, S., Hu, Z., 2007. Electricity Consumption and Economic Growth in China: Cointegration and Co-Feature Analysis. *Energy Economics* 29 (6), 1179-1191.
- Yusof, N., Latif, N., 2011. *Causality between Electricity Consumption and Economic Growth in Malaysia: Policy Implications*. Department of Economics & Entrepreneur Development College of Business Management & Accounting, Universiti Tenaga Nasional Sultan Haji Ahmad Shah Campus Pahang, Malaysia.
- Zachariadis, T., 2007. Exploring the Relationship between Energy Use and Economic Growth with Bivariate Models: New Evidence from G-7 Countries. *Energy Economics* 29 (6), 1233-1253.
- Zachariadis, T., Pashourtidou, N., 2007. An Empirical Analysis of Electricity Consumption in Cyprus. *Energy Economics* 29 (2), 183-198.
- Zamani, M., 2007. Energy Consumption and Economic Activities in Iran. *Energy Economics* 29 (6), 1135-1140.
- Zapata, H.O. and Rambaldi, A.N., 1997. Monte Carlo evidence on cointegration and causation, *Oxford Bulletin of Economics and Statistics* 59(2), 285-298.
- Zarnikau, J., 1997. A Reexamination of the Causal Relationship between Energy Consumption and Gross National Product. *The Journal of Energy and Development* 22 (1), 229-239.
- Zhang, X.-P., Cheng, X.-M., 2009. Energy Consumption, Carbon Emissions, and Economic Growth in China. *Ecological Economics* 68 (10), 2706-2712.
- Zhao, Z., Yuan, J., 2008. Income Growth, Energy Consumption and Carbon Emissions in China. *ICRMEM'08, The 2008 International Conference on Risk Management & Engineering Management*, IEEE Publications, pp. 373—377.
- Ziramba, E., 2009. Disaggregate Energy Consumption and Industrial Production in South Africa. *Energy Policy* 37 (6), 2214-2220.
- Zou, G., Chau, K., 2006. Short-and Long-Run Effects Between Oil Consumption and Economic Growth in China. *Energy Policy* 34 (18), 3644-3655.

Table 1. Studies Included in the Meta-Analysis

Authors and date	Countries	Aggregation of energy variable	Energy carriers	Aggregation of output	Control variables	Method
Abosedra and Baghestani (1991)	USA	M	TOT	M	-	G
Acaravici (2010)	TUR	M	EL	M	-	J
Adom (2011)	GHA	M	EL	M	-	TY
Akinlo (2008)	11 Sub-Saharan African countries	M	TOT	M	Energy pr., government exp.	G
Akinlo (2009)	NGA	M	EL	M	-	J
Alam et al. (2011)	IND	M	TOT	M	Empl., capital, energy pr., CO2	TY
Altinay and Karagol (2005)	TUR	M	EL	M	-	G
Ang (2008)	MYS	M	TOT	M	CO2	J
Belloumi (2009)	TUN	M	TOT	M	-	J
Boehm (2008)	15 EU countries	M	EL	M	-	G; J
Bowden and Payne (2009)	USA	M, C, I, R, T	TOT	M	Empl., capital	TY
Chiou-Wei et al. (2008)	8 Asian countries; USA	M	TOT	M	-	J
Chontanawat et al. (2008)	>100 countries	M	TOT	M	-	H
Ciarreta et al. (2009)	PRT	M	EL	M	Energy pr.	TY
Erol and Yu (1987)	CAN; FRA; ITA; JPN; GBR; DEU	M	TOT	M	-	G
Esso (2010)	CMR; COG; CIV; GHA; KEN;	M	TOT	M	-	TY
Fallahi (2011)	ZAF	M	TOT	M	-	G
Ghosh (2002)	USA	M	TOT	M	-	G
Glasure and Lee (1997)	IND	M	EL	M	-	G
Glasure (2002)	SGP; KOR	M	TOT	M	-	G; EG
Golam Ahamad and Nazrul Islam (2011)	KOR	M	TOT	M	Energy pr., gov. exp., money	J
Hondroyannis et al. (2002)	BGD	M	EL	M	-	J
Jamil and Ahmad (2010)	GRC	M, I, R	TOT	M	Energy pr.	J
Jamil and Ahmad (2011)	PAK	M, A, I, R, S	EL	M, A, I, R, S	Energy pr.	J
Jobert and Karanfil (2007)	POK	M, R	EL	M, R	Capital, Energy pr., degree days	J
Jumbe (2004)	TUR	M	TOT	M	-	G
Kaplan et al. (2011)	MWI	M	EL	M	-	EG; G
Karanfil (2008)	TUR	M	TOT	M	Empl., capital, energy pr.	J
	TUR	M	TOT	M	-	J

Notes: M: Macro; C: Commercial sector; T: Transport sector; I: Industry sector; R: Residential sector; S: Service sector; A: Agricultural sector

TOT: Total energy; TOTQA: Total quality adjusted energy; EL: Electricity; CO: Coal; O: Oil; NG: Natural Gas; F: Fuels; P: Petrol; RE: Renewables; N: Nuclear

G: Granger; S: Sims; EG: Engle-Granger; TY: Toda-Yamamoto; J: Johansen-Juselius; H: Hsiao

Table 1. Studies Included in the Meta-Analysis (continued)

Authors and date	Countries	Aggregation of energy variable	Energy carriers	Aggregation of output	Control variables	Method
Lee (2006)	G-11 countries	M	TOT	M	-	
Lorde et al. (2010)	BRB	M, Non-R	EL	M	Empl., capital, technology, CO2	J
Lotfalipour et al. (2010)	IRN	M	TOT, O, NG	M	CO2	TY
Masih and Masih (1996)	IND, PAK, MYS, SGP, PHL	M	TOT	M	-	J
Masih and Masih (1998)	THA, LKA	M	TOT	M	Energy pr.	J
Mehrara (2007)	IRN, KWT, SAU	M	TOT	M	-	TY
Menyah and Wolde-Rufael (2010a)	USA	M	N, RE	M	CO2	TY
Menyah and Wolde-Rufael (2010b)	ZAF	M	TOT	M	Capital, CO2	TY
Mozumder and Marathe (2007)	BGD	M	EL	M	-	J
Oh and Lee (2004)	KOR	M	TOTQA	M	Empl., capital	J
Pao and Tsai (2011)	RUS	M	TOT	M	CO2	J
Paul and Bhattacharya (2004)	IND	M	TOT	M	Capital, empl.	EG, H, J
Paul and Uddin (2010)	BGD	M	TOT	M	-	G
Payne (2009)	USA	M	RE, Non-RE	M	Empl., capital	TY
Payne (2010)	USA	M	N	M	Empl., capital	TY
Pradhan (2010)	BGD, IND, NPL, PAK, LKA	M	EL, O	M	-	J
Rafiq and Salim (2011)	IND, MYS, THA, CHN	M	TOT	M	Energy pr.	J
Sa.ad (2010)	NGA	M	TOT	M	-	J
Salim et al. (2008)	IND; CHN	M	TOT	M	Energy pr.	J
Sari and Soytaş (2009)	DZA, IND, NGA, SAU, VEN	M	TOT	M	Empl., CO2	TY
Shiu and Lam (2004)	CHN	M	EL	M	-	J
Soytas et al. (2007)	USA	M	TOT	M	Empl., capital, CO2	TY
Soytas and Sari (2009)	TUR	M	TOT	M	Empl., capital, CO2	TY
Türkekuş and Unakıtan (2011)	TUR	A	EL	A	Energy pr.	TY

Notes: M: Macro; C: Commercial sector; T: Transport sector; I: Industry sector; R: Residential sector; S: Service sector; A: Agricultural sector

TOT: Total energy; TOTQA: Total quality adjusted energy; EL: Electricity; CO: Coal; O: Oil; NG: Natural Gas; F: Fuels; P: Petrol; RE: Renewables; N: Nuclear

G: Granger; S: Sims; EG: Engle-Granger; TY: Toda-Yamamoto; J: Johansen-Juselius; H: Hsiao

Table 1. Studies Included in the Meta-Analysis (continued)

Authors and date	Countries	Aggregation of energy variable	Energy carriers	Aggregation of output	Control variables	Method
Vaona (2011)	ITA	M	RE; Non-RE	M	-	TY
Vecchione (2011)	ITA	M	EL	M	Degree days	J
Vlahinic-Dizdarevic and Zikovic (2010)	HRV	M; I; R	TOT	M	-	J
Wolde-Rufael (2009)	17 African countries	M	TOT	M	Employment.; capital	TY
Wolde-Rufael (2010a)	IND	M	N	M	Employment.; capital	TY
Wolde-Rufael (2010b)	CHN; IND; JPN; KOR; ZAF; USA	M	CO	M	Employment.; capital	TY
Wolde-Rufael and Menyah (2010)	9 developed countries	M	N	M	Employment.; capital	TY
Yoo (2006)	IND; MYS; SGP; THA	M	EL	M	-	H
Yoo and Ku (2009)	ARG; FRA; DEU; CHE	M	N	M	-	H
Yoo and Kwak (2010)	ARG; BRA; CHL; ECU; PER	M; C; I; R; T	EL	M	Empl.; capital	H
Yu and Hwang (1984)	USA	M	TOT	M	-	S
Yu and Choi (1985)	PHL; POL; KOR; GBR; USA	M	TOT; NG	M	-	G
Yuan et al. (2007)	CHN	M	EL	M	-	J
Yusuf and Latif (2011)	MYS	M	EL	M	-	G
Zachariadis (2007)	G7 countries	M; R; I; T; S	TOT	M; R; I; T; S	-	J; TY
Zachariadis and Pashourtidou (2007)	CYP	R; S	EL	R; S	Energy pr.; degree days	
Zamani (2007)	IRN	M; A; I	TOT; EL.; NG; P	M; A; I	-	J
Zhang and Cheng (2009)	CHN	M	TOT	M	Capital; CO2; popul.	TY
Zhao and Yuan (2008)	CHN	M	TOT	M	CO2	J
Ziramba (2009)	ZAF	M	CO; EL.; O	I	Employment.	TY
Zou and Chau (2006)	CHN	M	O	M	-	G; EG

Notes: M: Macro; C: Commercial sector; T: Transport sector; I: Industry sector; R: Residential sector; S: Service sector; A: Agricultural sector

TOT: Total energy; TOTQA: Total quality adjusted energy; EL: Electricity; CO: Coal; O: Oil; NG: Natural Gas; F: Fuels; P: Petrol; RE: Renewables; N: Nuclear

G: Granger; S: Sims; EG: Engle-Granger; TY: Toda-Yamamoto; J: Johansen-Juselius; H: Hsiao

Table 2. Distribution of Test Statistics

Energy Causes Output		
	Full Sample	Not Including Cointegration Tests
Sample mean	1.047	0.689
Standard deviation	1.577	1.394
Median	0.975	0.582
90 th Percentile	3.175	2.395
95 th Percentile	3.736	3.239
97.5 th Percentile	4.282	3.778
99 th Percentile	4.882	3.911
Sample Size	568	321

Output Causes Energy		
	Full Sample	Not Including Cointegration Tests
Sample mean	1.153	0.830
Standard deviation	1.646	1.448
Median	1.054	0.806
90 th Percentile	3.257	2.583
95 th Percentile	3.925	3.195
97.5 th Percentile	4.500	3.814
99 th Percentile	5.074	4.672
Sample Size	574	321

Table 3. Correlation Coefficients

	PREG	PRGE	RDFEG	RDFGE	SAMPLE
PREG	1.000	0.178	-0.055	-0.029	0.010
PRGE	0.178	1.000	0.016	-0.013	0.028
RDFEG	-0.055	0.016	1.000	0.942	0.833
RDFGE	-0.029	-0.013	0.942	1.000	0.822
SAMPLE	0.010	0.028	0.833	0.822	1.000
PUBYEAR	0.076	0.129	0.217	0.228	0.205
ZACHCI	0.182	0.222	0.281	0.274	0.155
START	0.043	0.052	-0.634	-0.622	-0.830
END	0.093	0.138	0.288	0.290	0.224
KEG	0.145	0.010	-0.587	-0.506	-0.094
KGE	0.098	0.064	-0.470	-0.581	-0.067
MACROE	-0.044	-0.059	-0.344	-0.333	-0.219
MACROG	0.000	-0.015	-0.301	-0.292	-0.162
MM	-0.034	-0.045	-0.337	-0.326	-0.203
FDIS	-0.181	-0.241	0.005	0.021	-0.071
CHIDIS	0.029	0.006	-0.074	-0.086	0.085
TDIS	0.202	0.240	0.094	0.086	0.009
LDC	-0.028	-0.033	-0.221	-0.198	-0.207
HDC	0.028	0.033	0.221	0.198	0.207
CONTROLS	0.150	0.076	-0.084	-0.097	0.086
VARIABLES	0.151	0.065	-0.113	-0.125	0.083
CI	0.259	0.217	0.152	0.136	0.011
TOTE	-0.068	-0.035	0.004	0.021	-0.047
TY	-0.078	-0.038	0.082	0.071	0.263
EMPLOYMENT	0.112	0.046	-0.191	-0.200	0.037
CAPITAL	0.129	0.079	-0.140	-0.148	0.086
PRICE	0.034	0.013	0.079	0.071	0.062
LAGSE_EG	0.100	-0.019	-0.475	-0.427	-0.062
LAGSG_EG	0.040	-0.069	-0.521	-0.415	-0.085
LAGSE_GE	0.031	-0.006	-0.394	-0.509	-0.048
LAGSG_GE	0.011	0.012	-0.401	-0.506	-0.047
TIME	0.150	0.070	-0.232	-0.236	-0.078
HSIAO	-0.108	-0.079	-0.180	-0.141	-0.210

EG – energy causes growth, GE – growth causes energy. For definitions of remaining variables see next page.

Approximate absolute critical values for a two tailed test for a sample of 574 observations:
 10%: ± 0.069 , 5%: ± 0.082 , 1%: ± 0.107 , 0.1%: ± 0.137

These are derived using: $\frac{r}{\sqrt{(1-r^2)/(N-2)}} \sim t(N-2)$

Variable Definitions:

PREG - -probit(p) statistic for energy causes growth
 PRGE - -probit(p) for growth causes energy
 RDFEG - square root of degrees of freedom for energy causes growth
 RDFGE - square root of degrees of freedom for growth causes energy
 SAMPLE - Original sample size before dropping any initial observations
 PUBYEAR - Year of publication
 ZACHCI - Dummy = 1 for cointegration results from Zachariadis (2007)
 START - First year of sample
 END - Last year of sample
 KEG - SAMPLE-DF for energy causes growth tests
 KGE - SAMPLE-DF for growth causes energy tests
 MACROE - Dummy = 1 if energy is measured at macro level
 MACROG - Dummy = 1 if output is measured at macro level
 MM - Dummy = 1 if both variables are measured at macro level
 FDIS - Dummy = 1 if original test statistic is an F-test
 CHIDIS - Dummy = 1 if original test statistic is Chi-squared
 TDIS - Dummy = 1 if original test statistic is t-test
 LDC - Dummy = 1 for less developed country
 HDC - Dummy = 1 for highly developed country
 CONTROLS - Number of control variables in model - e.g. for a model with energy, output, and capital this variable equals 1.
 VARIABLES - Number of controls + 2 or the number of time series in the VAR.
 CI - Dummy = 1 if model is cointegrated
 TOTE - Dummy = 1 if energy variable is total energy
 TY - Dummy = 1 if Toda-Yamamoto test was used
 EMPLOYMENT - Dummy = 1 if employment is included
 CAPITAL - Dummy = 1 if capital is included
 PRICE - Dummy = 1 if energy prices are included
 LAGSE_EG - Number of lags of energy in energy causes growth test
 LAGSG_EG - Number of lags of growth in energy causes growth test
 LAGSE_GE - Number of lags of energy in growth causes energy test
 LAGSG_GE - Number of lags of growth in growth causes energy test
 TIME - Dummy = 1 for model includes time trend
 HSIAO - Dummy = 1 for Hsiao procedure

Energy Causes Growth						Growth Causes Energy						
Model	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Constant	1.496 (4.04)	1.665 (4.37)	0.957 (4.39)	0.457 (1.98)	0.084 (0.28)	0.108 (0.36)	1.216 (3.32)	1.493 (3.94)	0.908 (3.89)	0.604 (3.30)	0.546 (1.11)	0.162 (0.48)
Square root DF	-0.487 (-3.70)	-0.517 (-4.09)					-0.291 (-2.19)	-0.340 (-2.58)				
DF			-0.010 (-1.78)						-0.003 (-0.41)			
CI	0.984 (4.51)	0.949 (4.14)	0.854 (3.76)	1.008 (4.61)	1.200 (5.30)	1.178 (5.00)	0.817 (2.93)	0.759 (2.63)	0.741 (2.58)	0.812 (2.87)	0.829 (2.55)	0.930 (2.90)
Sample	0.043 (2.75)	0.043 (3.01)		0.004 (0.76)	0.006 (0.85)	0.006 (0.91)	0.029 (2.60)	0.029 (2.72)		0.005 (1.10)	0.008 (1.72)	0.007 (1.45)
End		0.019 (1.57)						0.031 (2.37)				
KEG or KGE				0.069 (4.40)	-0.019 (-0.28)					0.036 (2.02)	0.092 (1.18)	
LAGSE					0.221 (1.71)	0.193 (2.83)					-0.157 (-1.19)	0.006 (0.21)
LAGSG					0.060 (0.47)	0.026 (0.57)					-0.046 (-0.38)	0.094 (1.34)
CONTROLSN UM					0.226 (1.33)	0.189 (1.90)					-0.096 (-0.72)	0.073 (0.71)
Time					0.742 (2.72)	0.700 (3.55)					0.208 (0.38)	0.426 (0.95)
HSIAO					0.398 (3.10)	0.382 (3.26)					0.243 (1.40)	0.308 (2.02)
Adjusted R- Squared	0.098	0.102	0.071	0.111	0.125	0.126	0.057	0.069	0.046	0.057	0.057	0.054
Jarque-Bera	0.55 (0.760)	0.54 (0.763)	1.60 (0.449)	0.96 (0.62)	0.36 (0.837)	0.45 (0.797)	48.85 (0.000)	45.39 (0.000)	46.79 (0.000)	48.77 (0.000)	47.90 (0.000)	49.08 (0.000)
Breusch- Pagan	0.42 (0.809)	0.44 (0.804)	0.07 (0.97)	0.07 (0.96)	0.22 (0.895)	0.23 (0.890)	2.20 (0.332)	2.31 (0.316)	2.32 (0.314)	2.32 (0.314)	2.59 (0.274)	2.28 (0.319)
t-stats in parentheses for regression coefficients, p-values for test statistics. Breusch-Pagan tests for heteroskedasticity related to the relevant degrees of freedom variable. This test statistic is chi-squared with 2 degrees of freedom.												

Table 8. Tests for Joint Effects of Methodologies and Variables

Energy Causes Growth	CI	SR	LR	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto	Hsiao
Named method & none of named variables	1.92	0.51	1.18	2.65	2.04	-1.58	0.10	-0.33	-1.93
Named method & other controls	1.61	0.48	0.81	2.35	1.63	n.a.	n.a.	-0.09	n.a.
Named method & capital	-1.30	1.21	-3.17	-0.47	-0.93	n.a.	n.a.	-0.44	n.a.
Named method & price	-0.05	-0.41	-0.46	1.09	-0.13	n.a.	0.97	-3.28	
Named method & time trend	1.91	0.51	1.66	3.55	1.81	0.13	1.06	-0.36	n.a.
Named method & employment	2.46	1.39	3.39	n.a.	2.60	n.a.	n.a.	0.25	n.a.
Named method & capital, employment, and time	1.39	1.36	1.04	n.a.	1.68	n.a.	n.a.	-0.48	n.a.

Figures are t-statistics. Bold indicates significant at the 5% level in a one-sided test.

Growth Causes Energy									
	CI	SR	LR	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto	Hsiao
Named method & none of named variables	1.41	-1.37	2.46	3.81	1.13	2.63	-1.69	0.46	-2.25
Named method & other controls	-0.33	-0.72	0.54	1.02	-0.35	n.a.	n.a.	-0.22	n.a.
Named method & capital	0.71	-1.92	-0.58	1.49	0.53	0.41	n.a.	1.06	n.a.
All other methods & price	1.81	3.83	1.31	1.31	1.98	2.74	2.85	2.96	2.47
Named method & price	3.44	0.62	3.69	3.91	3.51	n.a.	1.78	3.03	n.a.
Named method & time trend	1.05	-0.29	2.19	2.40	0.69	3.16	-0.06	-0.61	n.a.
Named method & employment	-0.59	-1.68	2.37	n.a.	-0.54	n.a.	n.a.	-2.65	n.a.

Figures are t-statistics. Bold indicates significant at the 5% level in a one-sided test.

Figure 1. Distribution of Energy Causes Growth Test Statistics Compared to Standard Normal Distribution

Figure 2. Distribution of Growth Causes Energy Test Statistics Compared to Standard Normal Distribution

Figure 3. Causal Structure

Figure 4. Distribution of Individual Test Statistics – Energy Causes Output

Figure 5. Distribution of Individual Test Statistics – Output Causes Energy