Is there really Granger Causality between Energy Use and Output?

Stephan B. Bruns, Christian Gross**, and David I. Stern****

ABSTRACT

We carry out a meta-analysis of the very large literature on testing for Granger causality 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 or misspecification bias. Our model extends the standard meta-regression model for detecting genuine effects in the presence of publication biases using the statistical power trace by controlling for the tendency to over-fit vector autoregression models in small samples. Granger causality tests in these over-fitted models have inflated type I errors. We cannot find a genuine causal effect in the literature as a whole. However, there is a robust genuine effect from output to energy use when energy prices are controlled for.

Keywords: Meta-analysis, Granger causality, Energy, Economic growth

http://dx.doi.org/10.5547/01956574.35.4.5

1. INTRODUCTION

The literature on Granger causality between energy and economic output consists of hundreds of papers. But despite attempts to review and organize this literature (e.g. Ozturk, 2010; Payne, 2010a), the nature of the relationship between the variables remains unclear (Stern, 2011). In this paper, we carry out a meta-analysis of this literature. Our goal is to determine whether there is a genuine causal relation between energy use and output or whether the large number of apparently significant results is due to publication or misspecification bias. 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.

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 (Card and Krueger, 1995). 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). Granger causality techniques have been widely applied in many areas of economics including monetary policy (Lee and Yang,

* Max Planck Institute of Economics, Kahlaischestr. 10, 07743 Jena, GERMANY. E-mail: stbruns@econ.mpg.de.
** Max Planck Institute of Economics, Kahlaischestr. 10, 07743 Jena, GERMANY. E-mail: gross@econ.mpg.de.
*** Corresponding author. Crawford School of Public Policy, Australian National University, Canberra, ACT 0200, Australia. E-mail: david.stern@anu.edu.au.

The Energy Journal, Vol. 35, No. 4.

2012), finance and economic development (Ang, 2008a), and energy economics (Ozturk, 2010), as well as in other fields 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; Payne, 2010b). In this paper, we show how meta-analysis can be used to test for genuine effects, publication, and misspecification biases in Granger-causality studies. The methods we use in this paper should be applicable to other areas of research that use Granger causality testing and possibly in the meta-analysis of studies using other econometric methods.

We modify the standard FAT-PET meta-regression model used in economics (Stanley and Doucouliagos, 2012) to meta-analyze Granger causality test statistics. The FAT-PET model regresses t-statistics from individual studies on the inverse of the standard errors of the regression coefficients of each study. If there is a genuine effect in the literature—a non-zero regression parameter—the coefficient of the inverse of the standard error will be non-zero, as the t-statistics will increase as the standard error declines in larger samples. This is the precision-effect test (PET). The intercept term is used to test for the presence of publication bias—the so-called funnel asymmetry test (FAT). Granger causality tests present three challenges to using the standard FAT-PET model. The first is that the usual restriction test statistics have an F or chi-squared distribution. These must be converted to statistics with a common distribution with properties that are suitable for regression analysis. We transform the p-values of the original test statistics to standard normal variates using the probit transformation.[1] The standard normal distribution is also better for meta-regression analysis than the commonly used t-distribution because the standard normal distribution is unaffected by the degrees of freedom. The second challenge is that these test statistics do not have associated standard errors. Therefore, our meta-regression model replaces the inverse of the standard error with the square root of the degrees of freedom of the regressions in the underlying studies. The third challenge is the tendency for researchers to over-fit vector autoregression (VAR) models in small samples (Gonzalo and Pitarakis, 2002). 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 (Zapata and Rambaldi, 1997). We control for these effects by including the number of degrees of freedom lost in fitting the underlying models as a control variable.

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 between energy and output 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 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. In this paper, we test for whether there are actual genuine effects in this literature rather than just misspecification and publication selection biases. Additionally, we use a larger sample consisting of 574 pairs of causality tests from 72 studies selected from this vast literature

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

of more than 500 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 Granger causality literatures. 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 a basic meta-regression analysis. 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 still find that there is no genuine effect in the literature as a whole but that models that include energy prices as a control variable have a genuine effect from output to energy use. Other effects are more fragile or ambiguous. The final section provides some suggestions and recommendations for future research.

2. METHODS

2.1 Testing for Genuine Effects

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 of interest—should have the same expected value across different studies irrespective of their degrees of freedom, *DF*. 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. Therefore, assuming for simplicity that the null hypothesis is $\beta = 0$, if there is a genuine non-zero effect, the absolute value of the related t-statistic should increase linearly with the square root of the degrees of freedom:

$$\begin{aligned} \frac{\hat{\beta}_i}{\hat{\sigma}_{\beta i}} &= t_i = \alpha DF_i^{0.5} + u_i \\ u_i &\sim t(DF) \end{aligned} \tag{1}$$

where i indexes individual test statistics[2] and α has the same sign as the underlying effect, β. The errors, u_i, are predictably heteroskedastic, as the variance of the t-distribution increases as the degrees of freedom decreases for low numbers of degrees of freedom. Card and Krueger (1995) and Stanley (2005a) suggest estimating a logarithmic version of (1), which Stanley calls meta-significance testing (MST):

$$\ln|t_i| = \ln \alpha_0 + \alpha_1 \ln DF_i + \varepsilon_i. \tag{2}$$

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, the heteroskedasticity of the t-distribution may introduce an undesirable negative correlation between the dependent variable and the degrees of freedom for low degrees of freedom. Second, due to taking absolute values and logarithms the error

2. 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.

term will not have a symmetric distribution, and will also be heteroskedastic if there is a genuine effect. Finally, though Stanley (2008) found (2) to be very powerful in large meta-samples even in the presence of publication biases, this test suffers from inflation of type 1 errors (Stanley, 2008; Stanley and Doucouliagos, 2012).

If all results are equally likely to be accepted for publication, there should be no relation between the estimated effect size and its standard error. However, if journals will only publish, or authors only submit for publication, statistically significant results, then, the lower the precision of estimation is, the larger reported effect sizes must be in order to achieve a given p-value and be published. This suggests a second meta-regression model:

$$\hat{\beta}_i = \gamma_0 + \gamma_1 \hat{\sigma}_{\beta i} + e_i \tag{3}$$

The test of $\gamma_1 = 0$, which Stanley (2005a) calls the funnel asymmetry test (FAT), 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. However, in the absence of publication bias, the variance of the errors in (3) will be correlated with the explanatory variable. So, Stanley (2005a) suggests that researchers divide both sides of (3) by the standard error and estimate the following model instead:

$$t_i = \gamma_0 \frac{1}{\hat{\sigma}_{\beta i}} + \gamma_1 + \upsilon_i \tag{4}$$

The same hypothesis tests apply to (4) as applied to (3) but it is now the intercept term that tests for publication bias and the slope coefficient is the estimate of the genuine effect. Stanley calls the test of $\gamma_0 = 0$ the precision effect test (PET). When we do not have information on standard errors, as in the case for most Granger causality tests, we can approximate the precision in (4) by the square root of the degrees of freedom (Stanley, 2005b):

$$t_i = \delta_0 + \delta_1 DF^{0.5} + \omega_i \tag{5}$$

But (5) is simply (1) with the addition of a constant. So, PET can be motivated by the same statistical power argument as we used to motivate MST (Stanley and Doucouliagos, 2012). But, unlike the MST model, (2), this model allows a neat decomposition of the sources of variance in the test statistics between the genuine effect, $\delta_1 DF^{0.5}$, and other sources of excess significance, δ_0. The errors in (5) will, in fact, still be heteroskedastic due to the nature of the t-distribution. This could be corrected by converting all test statistics to normal variates with the same significance levels as the t-statistics.

Up till this point, we have assumed that the effect size of interest is a regression coefficient, which was converted to a test statistic for the purpose of meta-analysis. However, in the case of Granger causality testing the effects of interest are the test statistics themselves. These are usually F or Chi-square distributed and in order to use them in a meta-regression they must be converted to normal variables. We convert them using the probit function—the inverse of the standard normal cumulative distribution. The transformation converts p-values of less than 0.5 into negative values and p-values greater than 0.5 into positive values. 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 non-causality at higher levels of significance:

$$-probit(p_i) = \alpha_0 + \alpha_1 DF^{0.5} + v_i \tag{6}$$

In the absence of publication bias, the intercept is expected to be zero, as $probit(0.5) = 0$. As only a positive relationship between the degrees of freedom and the dependent variable is associated with a genuine effect, we use a one-sided test of $H_0: \alpha_1 \leq 0$ in (6) to test for the presence of a genuine effect.

We give equal weight to each test statistic from each study 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 equations. In our initial estimates, in addition to our preferred model (6), we also estimate the MST model (2) and a version of (2) where we replace the t-statistics with normal variates.

2.2 Controlling for Overfitting

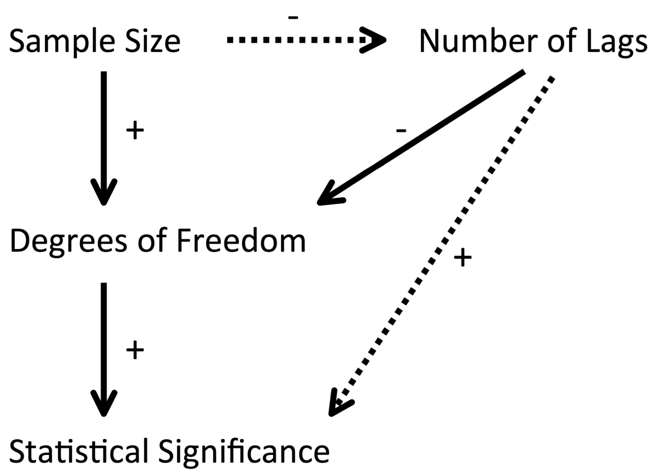
Researchers usually choose the number of lags of the variables in a VAR using the Akaike Information Criterion (AIC) or other goodness of fit indicators. The AIC, in particular, tends to over-estimate the number of lags when degrees of freedom are low and particularly when the VAR has a unit root or near unit root (Nickelsburg, 1985; Hacker & Hatemi-J, 2008; Gonzalo and Pitarakis, 2002). Zapata and Rambaldi (1997) show that three different causality tests over-reject the null hypothesis of non-causality when it is true in small samples especially when there is over-fitting. They assume that the data is I(1) and cointegrated with causality in only one direction, which allows comparable tests of both true and false Granger non-causality 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. They also find that the Toda-Yamamoto Granger causality 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 they also encourage researchers to add lags to the regression depleting degrees of freedom further. Figure 1 illustrates this causal structure assuming that there is a genuine effect. The solid channels are the statistical power relationship we want to estimate while the dashed channels are the over-fitting and over-rejection pathways that we want to exclude. In our sample, it appears that the dashed channels dominate 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 dashed path in Figure 1 removing the effects of intentional and unintentional data mining via model specification searches. Therefore we estimate:

$$-probit(p_i) = \alpha_0 + \alpha_1 DF_i^{0.5} + K_i + v_i \tag{7}$$

where K is the degrees of freedom lost in fitting the VAR in each study, which is composed of the number of coefficients estimated and the number of initial observations dropped due to adding lags of the variables. There is less of a problem with a large number of lags when the sample size is

Figure 1: Causal Structure



sufficiently large. We tried to take into account the possible reduced effect of over-fitting in larger samples by adding interaction terms in our empirical analysis but these had little effect and we do not report the results.

3. 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, 2010a) 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 included some unpublished studies in order to attempt to reduce publication bias. We completed our data collection in February 2012 and so we do not include any papers published after that date. We collected more than five hundred papers, but only a subset were coded and included in the meta-analysis. We filtered papers for commensurability and econometric quality and we 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 consumption of: total energy, 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. 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, 2012; Bruns and Gross, 2013). 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.

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	E prod, government exp.	G
Akinlo (2009)	NGA	M	EL	M	—	J
Alam et al. (2011)	IND	M	TOT	M	Empl, capital, E prod, CO2	TY
Altinay and Karagol (2005)	TUR	M	EL	M	—	G
Ang (2008b)	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	E prod	TY
Erol and Yu (1987)	CAN; FRA; ITA; JPN; GBR; DEU	M	TOT	M	—	G
Esso (2010)	CMR; COG; CIV; GHA; KEN; ZAF	M	TOT	M	—	TY
Fallahi (2011)	USA	M	TOT	M	—	G
Ghosh (2002)	IND	M	EL	M	—	G
Glasure and Lee (1997)	SGP; KOR	M	TOT	M	—	G; EG
Glasure (2002)	KOR	M	TOT	M	E prod, gov. exp., money	J
Golam Ahamad and Nazrul Islam (2011)	BGD	M	EL	M	—	J
Hondroyiannis et al. (2002)	GRC	M, I, R	TOT	M	E prod	J
Jamil and Ahmad (2010)	PAK	M, A, I, R, S	EL	M, A, I, R, S	E prod	J
Jamil and Ahmad (2011)	POK	M, R	EL	M, R	Capital, E prod, degree days	J
Jobert and Karanfil (2007)	TUR	M	TOT	M	—	G
Jumbe (2004)	MWI	M	EL	M	—	EG; G
Kaplan et al. (2011)	TUR	M	TOT	M	Empl, capital, E prod	J
Karanfil (2008)	TUR	M	TOT	M	—	J
Lee (2006)	G-11 countries	M	TOT	M	—	

(continued)

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
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	E prod	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 (2010a)	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	E prod	J
Sa'ad (2010)	NGA	M	TOT	M	—	J
Salim et al. (2008)	IND, CHN	M	TOT	M	E prod	J
Sari and Soytas (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
Türkekul and Unakıtan (2011)	TUR	A	EL	A	E prod	TY
Vaona (2012)	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 and Menyah (2010)	9 developed countries	M	N	M	Empl, capital	TY
Wolde-Rufael (2009)	17 African countries	M	TOT	M	Empl, capital	TY
Wolde-Rufael (2010a)	IND	M	N	M	Empl, capital	TY

(continued)

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
Wolde-Rufael (2010b)	CHN, IND, JPN, KOR, ZAF, USA	M	CO	M	Empl, 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
Yusof 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	E prod, 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, population	TY
Zhao and Yuan (2008)	CHN	M	TOT	M	CO2	J
Ziramba (2009)	ZAF	M	CO, EL, O	I	Empl	TY
Zou and Chau (2006)	CHN	M	O	M	—	G, EG

Abbreviations:
Aggregation: M: Macro; C: Commercial sector; T: Transport sector; I: Industry sector; NR: Non-residential sector; R: Residential sector; S: Service sector; A: Agricultural sector
Energy: 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
Controls: Empl: employment; E prod: energy production
Method: G: Granger; S: Sims; EG: Engle-Granger; TY: Toda-Yamamoto; J: Johansen-Juselius; H: Hsiao

We included 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 only, or a joint short and long run test. We excluded models that include contemporaneous terms on the right hand side of the regression 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 including nonparametric approaches (e.g. Azomahou et al., 2006), threshold cointegration (e.g. Esso, 2010), and those including structural breaks.[3] For reasons of comparability we also excluded all studies that found more than one cointegrating vector using the Johansen approach, such as Stern (2000).

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 so that we could test the effects of the level of economic development. Similarly, we excluded studies for the sub-country level, e.g., cities, regions, and provinces, including Taiwan, for reasons of comparability.

We could only include those studies that include all necessary information, 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 and if we did not receive any reply or if the answer was still incomplete, we excluded those studies.

We excluded studies with incorrect estimation strategies such as using VARs with different lag lengths for the Johansen-Juselius cointegration test and for the VECM based on the estimated cointegration vectors and, importantly, all Granger causality tests on VARs in levels that do not use the Toda-Yamamoto approach. We thus excluded a large number of early studies including Stern (1993). Finally, we excluded unclear or statistically incorrect presentations of results such as negative F statistics or studies that only reported significant results (e.g. Chang and Soruco-Carballo, 2011). The aim of these exclusions was to reduce the effect of spurious regression or other econometric errors on the meta-analysis. We documented the reasons for excluding all excluded studies. This information is available on request.

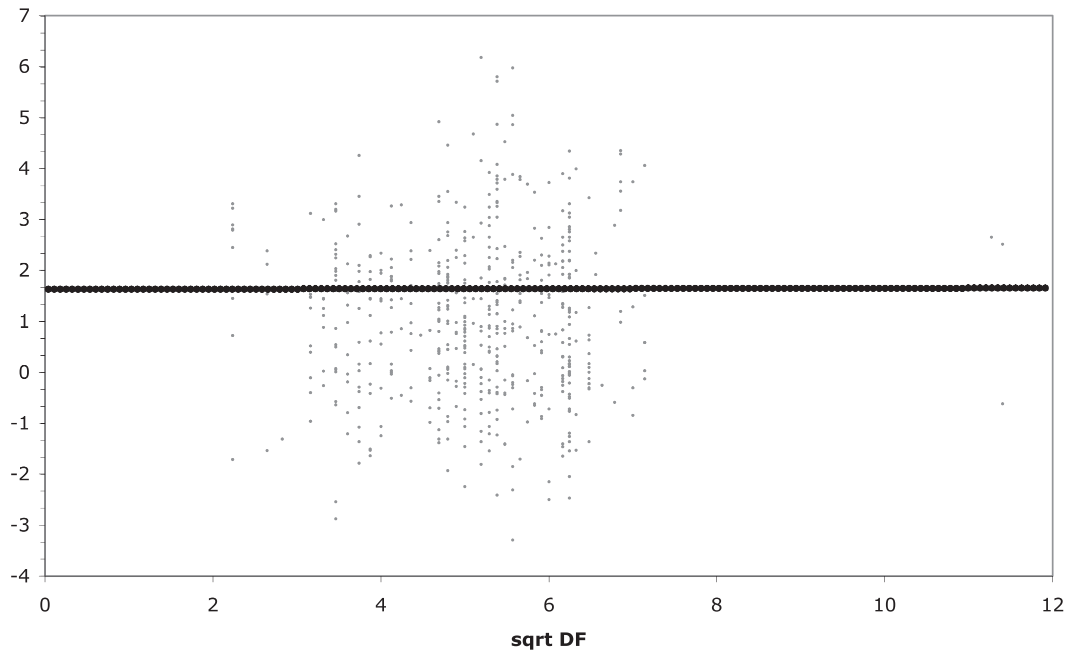
4. EXPLORATORY ANALYSIS OF THE DATA

4.1 Description of the Data

The full sample comprises 72 studies. We have 574 growth causes energy test statistics and 568 energy causes growth test statistics. There are a total of 428 macro-macro only observations (425 in the energy to 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 to GDP direction of causality). Following current Intergovernmental Panel on Climate Change practice, we

3. For Esso (2010), we coded the Toda-Yamamoto causality tests included in the paper and similarly for Vaona (2012) we coded observations that did not use structural breaks. There are many such instances where we only partially coded a paper.

Figure 2: Distribution of Individual Test Statistics by Degrees of Freedom—Energy Causes Output



regard countries as OECD countries if they were members of the OECD in 1990. 264 observations are from countries that were members of the OECD in 1990 and 310 from other countries.

In all cases, we assume that the sample size associated with each test statistic is the length of the time series used despite the fact that some test statistics were produced using 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 only used it to estimate the cointegrating vectors. They then estimate a VECM using OLS with predetermined error correction terms derived from the Johansen estimates and make inferences from these second stage estimates.

Figures 2 and 3 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 in each figure is for a test statistic value of 1.65, which is the critical value at the 5% significance level. The outliers to the right in each figure are from Vaona (2012)—the study with the longest time series in our dataset. There does not seem to be a strong relationship between degrees of freedom and the values of the test statistics.

Table 2 provides information on the distribution of the probit transformed p-values for the full sample and for a sub-sample excluding the cointegration studies. The distributions for OECD and non-OECD sub-samples (not reported) are very similar. For the full sample, 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 statistic is not significant at conventional levels. On the other hand, these means are much greater than the zero value expected under the null hypothesis of non-causality.[4] Additionally, the standard deviations in Table 2 are greater than

4. The t-statistics for the difference of the means from zero are 15.8 and 16.8 respectively.

Figure 3: Distribution of Individual Test Statistics by Degrees of Freedom—Output Causes Energy

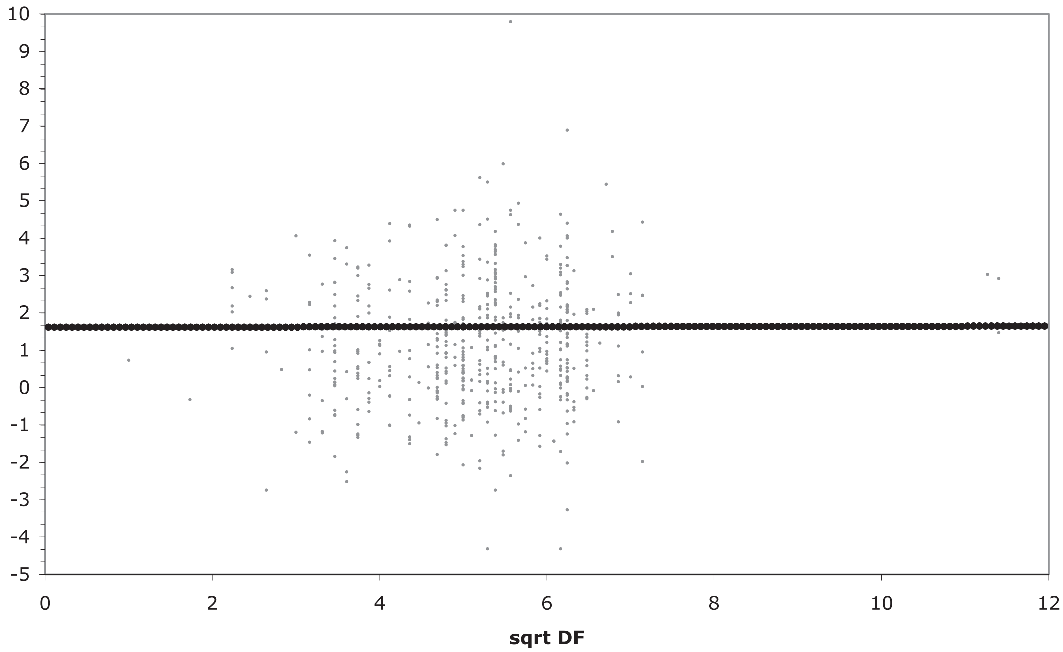


Table 2: Distribution of the Probit Transformed p-Values

Energy Causes Output

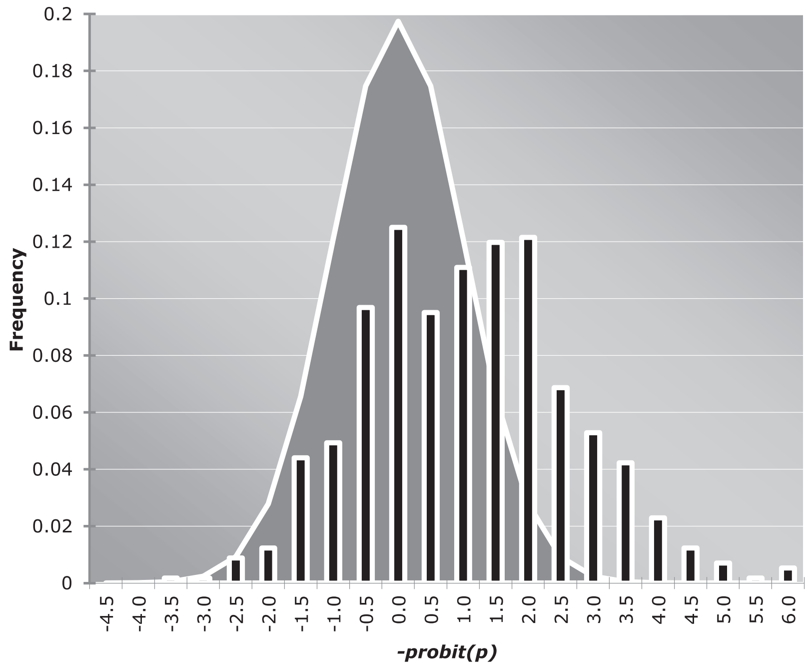
	Full Sample	Not Including Cointegration Tests
Sample mean	1.047	0.689
Standard deviation	1.577	1.394
Median	0.975	0.582
90th Percentile	3.175	2.395
95th Percentile	3.736	3.239
97.5th Percentile	4.282	3.778
99th Percentile	4.882	3.911
Sample Size	568	321

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
90th Percentile	3.257	2.583
95th Percentile	3.925	3.195
97.5th Percentile	4.500	3.814
99th Percentile	5.074	4.672
Sample Size	574	321

Notes: As explained in the text, the probit values are multiplied by −1 so that larger numbers indicate more statistically significant test statistics.

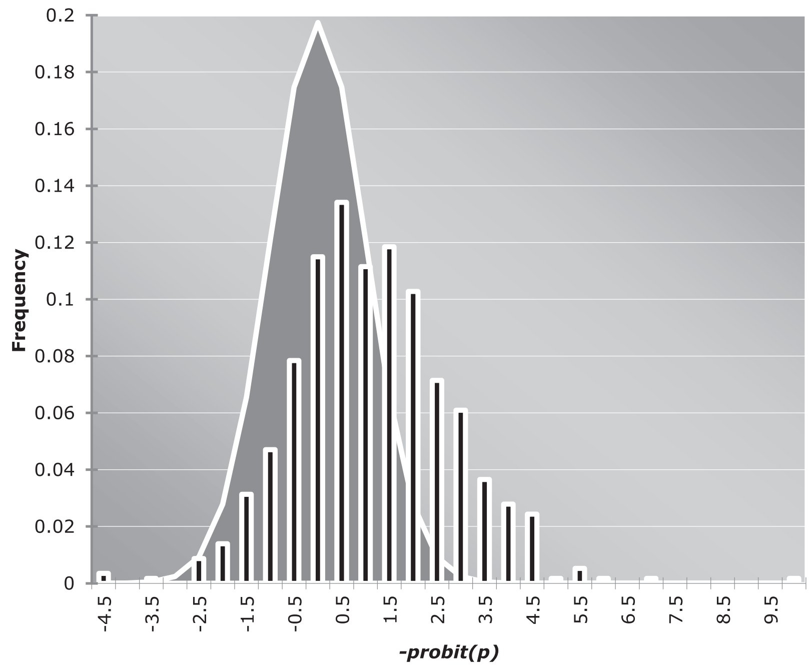
Figure 4: Distribution of Energy Causes Growth Test Statistics Compared to Standard Normal Distribution



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 listed 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. Figures 4 and 5 compare the distribution of the test statistics to the standard normal distribution. Both samples lower central frequencies than the standard normal distribution with a flat top to the distribution curve and a fat upper tail. The deviation of these distributions from the standard normal 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 conventional levels and there is no apparent relationship between the test statistics and degrees of freedom in Figures 2 and 3.
b. Publication bias results in many statistically insignificant results not being reported and/or authors carry out specification searches to generate more significant test statistics.
c. Spurious regression might produce seemingly significant results. 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. If two variables cointegrate, then there must be Granger causality between them in at least one direction (Engle and Granger, 1987). Prescreening for cointegration and often

Figure 5: Distribution of Growth Causes Energy Test Statistics Compared to Standard Normal Distribution



inappropriate methods of testing for causality in cointegrated models may exaggerate reported significance levels (Clarke and Mirza, 2006). To test this explanation, we also present in Table 2 statistics for samples excluding the results of cointegration tests. Though this lowers average levels of significance, there is still a lot of excess significance that needs to be explained. We will test explanations a., b., and c. in the remainder of the paper while controlling for the effect of cointegration pre-screening.

4.2 Correlation Analysis

Table 3 presents the correlations between the main variables of interest and all other variables. A key to the variable names follows the table. The correlations are fairly similar for the causality tests in each direction. The correlations between the test statistics (PREG and PRGE) and the square root of degrees of freedom (RDFEG and RDFGE) are negative but weak (-0.055 and -0.013). This is the opposite of the expected relationship if there were genuine effects. There are very weak positive relations between the test statistics and sample size. But the number of degrees of freedom lost in fitting the regressions (KEG and KGE) is positively associated with the test statistics (significant at the 0.1% level for energy causes growth and at 5% for growth causes energy). As noted above, the test statistics are significantly higher when a cointegration test has been passed (CI). Models that include CAPITAL (which includes gross fixed capital formation as well as the capital stock) or EMPLOYMENT are also more significant. Later sample START dates are positively but weakly associated with the test statistics as are later sample END dates and the PUBLICATION YEAR. So it seems that the relationship between energy and growth may have

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
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
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
CONTROLS	0.150	0.076	−0.084	−0.097	0.086
VARIABLES	0.151	0.065	−0.113	−0.125	0.083
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
OTHER	0.117	0.028	−0.031	−0.038	0.011
TIME	0.150	0.070	−0.232	−0.236	−0.078
CI	0.259	0.217	0.152	0.136	0.011
TY	−0.078	−0.038	0.082	0.071	0.263
HSIAO	−0.108	−0.079	−0.180	−0.141	−0.210
TOTE	−0.068	−0.035	0.004	0.021	−0.047
MM	−0.034	−0.045	−0.337	−0.326	−0.203
OECD	−0.002	0.034	0.389	0.375	0.329
non-OECD	0.002	−0.034	−0.389	−0.375	−0.329

Notes: 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: $\frac{r}{\sqrt{(1-r^2)/(N-2)}} \sim t(N-2)$

Key to Variable Definitions in Table 3:

Some variables have different values for the tests in each direction. These are treated as two different variables. We label these pairs of variables with the suffix EG for "Energy causes Growth" and GE for "Growth causes Energy". All other variables have a common value for both tests.

PREG and PRGE: *probit*(p) statistic
RDFEG and RDFGE : square root of degrees of freedom
SAMPLE: Original time series sample size before dropping any initial observations
START: First year of the sample
END: Last year of the sample
KEG and KGE: Degrees of freedom lost in fitting the regression equation
LAGSE_EG and LAGSE_GE: Number of lags of energy
LAGSG_EG and LAGSG_GE: Number of lags of output
CONTROLS: Number of control variables in model—e.g. for a model with energy, output, and capital this variable equals 1
VARIABLES: Number of control variables + 2 or the number of time series in the VAR
EMPLOYMENT: Dummy = 1 if employment is included
CAPITAL: Dummy = 1 if capital is included
PRICE: Dummy = 1 if energy prices are included
OTHER: Controls other than employment, capital, and energy price
TIME: Dummy = 1 for model includes time trend
CI: Dummy = 1 if model is cointegrated
TY: Dummy = 1 if Toda-Yamamoto test was used
HSIAO: Dummy = 1 for Hsiao procedure
TOTE—Dummy = 1 if energy variable is total energy
MM: Dummy = 1 if both energy and output are measured at macro level
OECD: Dummy = 1 for countries that were members of the OECD in 1990
nonOECD: Dummy = 1 for countries that were not members of the OECD in 1990

strengthened over time though of course this does not control for changes in methodology and in the sample of countries.

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 KEG and KGE. At first glance, this appears to make sense—increasing the number of variables and lags reduces the degrees of freedom. But that is only true holding the sample size constant! We should expect that, as the sample size grows, both the number of variables and the degrees of freedom would increase if researchers add extra variables at a slower rate than the sample size is increasing. Sample size is also somewhat negatively correlated with KEG and KGE. These correlations support explanation c. above, suggesting that there is a tendency to over-fit models in small samples and for those models to have inflated type I errors. There is also a negative correlation between the sample size and the number of lags (LAGSE and LAGSG) 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 PRICE. Finally, the number of degrees of freedom and sample sizes are greater for the OECD countries than the developing countries but correlations between the test statistics and development status are not statistically significant.

4.3 Basic Meta-Regression Analysis

Table 4 presents the results for the basic meta-regression models. The first three columns for both directions of causation are the simple meta-regression models (2) or (6) while the second three columns add a control for cointegration. Results are remarkably similar for tests of both directions of causality except that the energy to growth direction is generally slightly more significant (higher R-squared). The degrees of freedom variable always has a negative coefficient, but the negative slope is reduced by using normal variates instead of t-statistics and is not significantly less than zero for three out of four of the probit transform models. As explained above, though we might expect a negative slope when the dependent variable is the log of the t-statistic, it is unexpected to find a negative slope for the other two forms of the dependent variable.

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. Screening for cointegration should result in significant Granger causality in at least one direction. On average, it is found in both directions.

The residuals would be expected to be heteroskedastic for the logarithmic models in the presence of a genuine effect and whether there is or is not a genuine effect when the t-statistics are used. We cannot reject the null hypothesis of homoskedasticity for any of the models at the 5% level. However, the residuals of 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. The intercept terms of the probit transform models are highly significant, suggesting publication or misspecification bias.

The Chow tests show that we cannot reject pooling the data for developing and developed countries into a single dataset for the regressions that do not control for cointegration. However, there is a significant difference between the two groups for the regressions that do control for cointegration. The main difference is that the coefficient of degrees of freedom is more negative in developing countries than developed countries. While there is a difference, the difference is not

Table 4: Basic Meta-Significance Testing

	Energy Causes Growth						Growth Causes Energy					
Dependent variable	$\ln t_{EG}$	$\ln Z_{EG}$	PREG	$\ln t_{EG}$	$\ln Z_{EG}$	PREG	$\ln t_{GE}$	$\ln Z_{GE}$	PRGE	$\ln t_{GE}$	$\ln Z_{GE}$	PRGE
Constant	0.969	0.651	1.436	1.025	0.704	1.351	1.013	0.640	1.209	1.051	0.674	1.123
	(2.99)	(2.17)	(5.15)	(2.67)	(1.97)	(3.62)	(2.89)	(2.17)	(3.40)	(2.95)	(2.16)	(2.73)
Log DF	−0.272	−0.192		−0.358	−0.273		−0.269	−0.174		−0.343	−0.243	
	(−2.65)	(−2.02)		(−3.21)	(−2.62)		(−2.53)	(−1.92)		(−3.15)	(−2.48)	
Sqrt DF			−0.076			−0.134			−0.011			−0.058
			(−1.49)			(−2.10)			(−0.17)			(−0.79)
CI				0.504	0.476	0.871				0.459	0.430	0.753
				(3.09)	(2.99)	(3.89)				(2.09)	(2.02)	(2.64)
Adjusted R-Squared	0.009	0.004	0.001	0.046	0.039	0.073	0.007	0.002	−0.002	0.031	0.025	0.047
Skewness	−1.60	−1.77	0.25	−1.78	−1.96	0.11	−3.02	−3.33	0.37	−3.07	−3.37	0.26
	(0.000)	(0.000)	(0.015)	(0.000)	(0.000)	(0.304)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.012)
Kurtosis	4.37	5.00	−0.04	5.54	6.23	0.15	18.33	20.71	1.35	18.28	20.63	1.30
	(0.000)	(0.000)	(0.837)	(0.000)	(0.000)	(0.47)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Jarque-Bera	694.5	889.7	6.02	1028.0	1282.8	1.58	8912.8	11318.6	56.45	8893.1	11270.5	46.83
	(0.000)	(0.000)	(0.049)	(0.000)	(0.000)	(0.453)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Breusch-Pagan	1.00	1.69	1.34	0.51	1.01	0.10	2.20	2.65	5.21	1.86	2.28	2.66
	(0.607)	(0.429)	(0.512)	(0.774)	(0.603)	(0.950)	(0.332)	(0.266)	(0.074)	(0.39)	(0.32)	(0.264)
OECD	1.399	1.145	0.711	4.227	4.094	4.878	0.357	0.421	0.716	2.395	2.328	4.278
	(0.248)	(0.319)	(0.492)	(0.006)	(0.007)	(0.002)	(0.700)	(0.657)	(0.489)	(0.067)	(0.073)	(0.005)

Notes: 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 distributed as chi-squared with 2 degrees of freedom. OECD is an F-test for pooling of OECD and non-OECD countries. Dependent variables: lnt—the log of the positive two-sided t-statistic corresponding to the original p-value. lnZ—the log of the positive two-sided normal variate corresponding to the original p-value. PREG and PRGE are as in Table 3.

radical. Results were similar when we defined developed countries using the current World Bank income based definition.

5. EXPLAINING THE NEGATIVE SLOPE

5.1 Alternative Hypotheses

We think that the most likely explanation for the negative effect of degrees of freedom in Table 4 is the over-fitting, over-rejection hypothesis. We examine this first and then examine some alternative hypotheses. 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. There are also negative correlations between the sample size and the number of lags and the presence of a time trend, but a positive correlation between sample size and the number 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 is positively correlated with the test statistics for those tests though this correlation is not significant at the 10% level.

We explore this potential effect using the regressions reported in Table 5.[5] Regressions II to IV decompose degrees of freedom into its two components—the original sample size and the number of degrees of freedom lost in model fitting, *K*. Regression I in Table 5 checks whether using a linear rather than a square root function of degrees of freedom significantly affects the results. It does not. The results are very similar to the corresponding results in Table 4.

The coefficient of SAMPLE in regressions II (Table 5) shows the effect of increasing the sample size while holding *K* constant. Sample size does not have a significant effect on the dependent variable, *ceteris paribus*, while *K* has a significant positive effect. This result strongly supports our hypothesis, as an increase in *K* holding the sample size constant is equivalent to a decrease in the degrees of freedom.

In regression IV, we replace *K* with the various variables that can be included in the underlying VAR models. We also add a dummy for the Hsiao procedure because this approach results in a different numbers of lags for each variable in each equation. 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 at the 10% level 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 each variable individually to deliberately get the most significant fit. We also tried dropping one of the lags variables from regression IV but this made little difference to the results.

5. The Chow statistics show that results in Table 5 differ for developing and developed countries. The developed country regressions have better goodness of fit and generally more significant regression coefficients. Significant coefficients have the same signs in both sets of regressions. So, while the differences are significant they are again not radical.

Table 5: Alternative Explanations of the Negative Slope Coefficient

	Energy Causes Growth				Growth Causes Energy			
Model	I	II	III	IV	I	II	III	IV
Constant	0.957	0.457	0.554	0.108	0.908	0.604	0.772	0.162
	(4.39)	(1.98)	(2.488)	(0.36)	(3.89)	(3.30)	(3.41)	(0.48)
DF	−0.010				−0.003			
	(−1.78)				(−0.41)			
CI	0.854	1.008	0.974	1.178	0.741	0.812	0.752	0.930
	(3.76)	(4.61)	(4.249)	(5.00)	(2.58)	(2.87)	(2.56)	(2.90)
Sample		0.004	0.002	0.006		0.005	0.001	0.007
		(0.76)	(0.38)	(0.91)		(1.10)	(0.23)	(1.45)
End			0.017				0.29	
			(1.43)				(2.21)	
K		0.069	0.071			0.036	0.039	
		(4.40)	(4.66)			(2.02)	(2.29)	
LAGSE				0.193				0.006
				(2.83)				(0.21)
LAGSG				0.026				0.094
				(0.57)				(1.34)
CONTROLSNUM				0.189				0.073
				(1.90)				(0.71)
Time				0.700				0.426
				(3.55)				(0.95)
HSIAO				0.382				0.308
				(3.26)				(2.02)
Adjusted R^2	0.071	0.111	0.114	0.126	0.046	0.057	0.068	0.054
Jarque-Bera	1.60	0.96	0.835	0.45	46.79	48.77	46.01	49.08
	(0.449)	(0.62)	(0.658)	(0.797)	(0.000)	(0.000)	(0.000)	(0.000)
Breusch-Pagan	0.07	0.07	0.10	0.23	2.32	2.32	2.93	2.28
	(0.97)	(0.96)	(0.949)	(0.890)	(0.314)	(0.314)	(0.231)	(0.319)
OECD	5.318	4.490	3.304	2.282	4.020	4.584	3.972	3.347
	(0.001)	(0.001)	(0.006)	(0.021)	(0.008)	(0.001)	(0.002)	(0.001)

Notes: 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 distributed as chi-squared with 2 degrees of freedom. OECD is an F-test for pooling of OECD and non-OECD countries. See Table 3 for variable definitions.

From this it is clear that 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 tested this hypothesis by running regression II using only those studies that report a single sample size.[6] 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 positively associated with sample size. This suggests that control variables are not added to re-

6. 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.

gressions 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 accords with Gonzalo and Pitarakis' (2002) finding that over-fitting is less likely in higher dimensional VARs.

In regressions II to IV the variables for coefficient, lags, and control numbers are demeaned, so that the intercept term is for a model with average numbers of these. The intercept is much reduced for these three models and for regression IV insignificant, suggesting that much of the excess significance in the simple meta-regression model I is in fact due to misspecification bias.

We also examine three alternative hypotheses that might explain the negative coefficient on degrees of freedom in Table 4:

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. However, the relatively low correlation between end date and sample size shows that larger samples are not strongly associated with more recent data and the positive relation between both start and end dates and the value of the test statistics (Table 3) shows that more recent data is likely to have higher test statistics. We can also test this hypothesis with regression III in Table 5. Holding the sample size constant and increasing the end date, effectively moves a time window of fixed length through the data as the start date is implicitly increased in line with the end date. 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 allows us to reject the hypothesis that the test statistics are smaller in more recent.
2. There may be more structural 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. Sample size has a positive but insignificant effect on the dependent variable in all the regressions in which it is included in Table 5. Also, from Table 3 we see that sample size has a very weak positive simple correlation with the test statistics. Therefore, this hypothesis cannot explain the negative slope of degrees of freedom though it could be a reason why sample size has only weak effects on the test statistic.
3. Authors with smaller samples could be more prone to trying to get significant results than authors with larger samples. Then, if there were no genuine effect the slope of degrees of freedom would be negative. Though this is possible, it is not testable.

The evidence, therefore, strongly supports the over-fitting, over-rejection hypothesis as the explanation of the negative slope of the degrees of freedom variable that we found in Table 4. Table 6 presents estimates of versions of the meta-Granger causality model (7). Degrees of freedom have a positive but insignificant coefficient in five of the six regressions in the table. Therefore, we conclude that there is no observable genuine effect in the meta-sample as a whole. The effect is stronger in the models that control for total coefficients and dropped initial observations (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. Differences between developing and developed countries are less significant for these regressions than they are for the regressions in Table 5.

Table 6: Meta-Granger Causality Tests

	Energy Causes Growth			Growth Causes Energy		
Model	A	B	C	A	B	C
Constant	0.292	0.625	0.0002	0.545	0.583	0.283
	(0.79)	(1.72)	(0.00)	(1.66)	(1.42)	(0.69)
Sqrt DF	0.062	−0.007	0.094	0.049	0.002	0.078
	(0.91)	(−0.12)	(1.20)	(0.76)	(0.02)	(1.11)
KEG or KGE	0.076		0.071	0.041		0.042
	(3.81)		(3.03)	(2.08)		(1.89)
CI	1.007	1.055	1.081	0.810	0.825	0.898
	(4.60)	(4.73)	(4.67)	(2.86)	(2.84)	(3.03)
HSIAO			0.255			0.258
			(1.84)			(1.68)
LAGSE_EG or LAGSG_GE		0.229			0.098	
		(3.18)			(1.52)	
TIME			0.645			0.330
			(3.20)			(0.72)
Adjusted R-Squared	0.111	0.103	0.120	0.057	0.051	0.058
Jarque-Bera	1.08	0.80	0.68	48.63	50.63	48.92
	(0.58)	(0.67)	(0.71)	(0.00)	(0.00)	(0.00)
Breusch-Pagan	0.22	0.00	0.16	2.58	2.24	2.62
	(0.89)	(0.99)	(0.92)	(0.27)	(0.32)	(0.27)
OECD	3.263	3.399	2.240	2.094	1.920	1.904
	(0.012)	(0.009)	(0.038)	(0.080)	(0.106)	(0.078)

Notes: 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 distributed as chi-squared with 2 degrees of freedom. OECD is an F-test for pooling of OECD and non-OECD countries. See Table 3 for variable definitions.

6. EFFECTS OF METHODOLOGY ON FINDING A GENUINE EFFECT

Ozturk (2010) and Stern (2011) both argue that some methods are more likely to uncover a robust effect than others. Though we cannot find a genuine Granger causality effect in the sample as a whole, perhaps some methodological approaches do uncover real causality effects. In this section we test for whether there are any methodologies where we can find a genuine effect. The methodologies include both econometric methods and the inclusion of various control variables. We test their effects by adding a dummy variable and an interaction term between the dummy and the degrees of freedom variable to our meta-regression model:

$$-probit(p_i) = \alpha_0 + \alpha_1 DF_i^{0.5} + \alpha_2 K_i + \beta_0 d_i + \beta_1 d_i DF_i^{0.5} + v_i \tag{8}$$

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-statistics for $\alpha_1 + \beta_1$ and $\alpha_0 + \beta_0$ 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 due to publication or misspecification bias 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.

Where we do find genuine effects these indicate that GDP causes energy. There do appear to be genuine effects for cointegrated results. The result in the growth causes energy direction is

Table 7: Tests for Effects of Methodologies, Variables, and Variable Definitions

	Techniques								
Methodology	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto	Hsiao
Energy Causes Growth									
Levels coefficient	0.668	0.609	1.373	0.114	0.536	1.680	0.162	0.959	1.574
	(0.91)	(1.06)	(1.47)	(0.06)	(0.66)	(4.78)	(0.22)	(1.69)	(2.51)
Joint slope coefficient	0.180	0.099	0.095	0.321	0.200	−0.007	0.064	−0.068	−0.187
	(1.16)	(0.91)	(0.52)	(0.86)	(1.20)	(−0.09)	(0.44)	(−0.72)	(−1.37)
Growth Causes Energy									
Levels coefficient	−0.076	1.922	−0.063	−4.400	0.103	−1.718	1.369	1.366	1.830
	(−0.08)	(2.07)	(−0.08)	(−2.77)	(0.10)	(−1.93)	(2.19)	(1.56)	(2.36)
Joint slope coefficient	0.315	−0.248	0.459	1.137	0.283	0.649	−0.188	−0.088	−0.195
	(1.85)	(−1.46)	(3.23)	(3.97)	(1.56)	(3.76)	(−1.44)	(−0.56)	(−1.20)

	Variables				Variable Definition	
Methodology	Time	Price	Capital	Employment	Macro-macro	MM Total Energy
Energy Causes Growth						
Levels coefficient	0.818	1.545	2.174	1.326	0.879	0.360
	(0.75)	(1.83)	(2.16)	(1.36)	(1.83)	(0.58)
Joint slope coefficient	0.190	−0.061	−0.172	−0.020	0.018	0.107
	(0.77)	(−0.40)	(−0.92)	(−0.11)	(0.20)	(0.88)
Growth Causes Energy						
Levels coefficient	2.545	−1.325	2.939	2.461	0.559	0.509
	(1.92)	(−1.39)	(3.02)	(2.25)	(1.70)	(1.08)
Joint slope coefficient	−0.267	0.478	−0.305	−0.278	0.104	0.112
	(−0.78)	(3.15)	(−1.81)	(−1.38)	(1.60)	(1.17)

Notes: t-stats in parentheses.

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. 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 control variables, only those models that include energy prices have a genuine effect for GDP causes energy. This model defines a demand function where energy use is determined by prices and income. Note that energy might still cause income when changes in energy use are driven by changes in energy prices. 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-value = 0.054). But 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. We also tested for effects in a sample that excludes the cointegration studies. These results were also similar to those in Table 7. We also carried out all the tests in Table 7 for OECD and non-OECD countries separately (reported in the Appendix). Results were again fairly similar for the two groups.

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 had results that were much different to those reported in Table 7 in terms of sign or significance level.

Finally, we tested joint hypothesis of whether there are genuine effects when using particular control variables with specific methods by estimating the following regression for each method:

$$-probit(p_i) = \alpha_0 + \alpha_1 DF_i^{0.5} + \alpha_2 K_i + \beta_0 m_i + \beta_1 m_i DF_i^{0.5} + \sum_j (\gamma_{0j} c_{ji} + \gamma_{1j} c_{ji} DF_i^{0.5} + \gamma_{2j} m_i c_{ji} DF_i^{0.5}) + v_i \quad (9)$$

where m is the dummy variable for the method under investigation and the c_j are dummies for the various possible control variables in the underlying studies. The interaction terms between the method and control variable dummies and the square root of degrees of freedom test 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 dummy variables alone, but we found these effects to be insignificant and dropped them. This model is estimated separately for each method.

We report the results in Table 8 in terms of t-statistics for linear combinations of regression coefficients that measure the stated treatments. We dropped the Hsiao method, as it does not use control variables. The first row of Table 8 is a test of $\alpha_1 + \beta_1 > 0$, which is a test of a genuine effect when the named method is used but no control variables are included. The following rows test $\alpha_1 + \beta_1 + \gamma_1 + \gamma_2 > 0$, which is a test of whether there is a genuine effect when this method and control is used and other controls are set to zero.

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

Table 8: Tests for Joint Effects of Methodologies and Variables

Energy Causes Growth

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	**2.21**	0.67	1.26	**1.96**	**2.32**	−1.37	0.40	−0.11
Capital	−1.54	−1.30	−3.28	−0.87	−1.15	n.a.	n.a.	−0.50
Price	0.10	−0.38	−0.40	0.59	0.03	n.a.	1.23	−3.36
Time trend	**2.00**	0.56	**1.67**	**3.00**	**1.77**	0.36	1.22	−0.37
Employment	**2.66**	1.57	**3.59**	n.a.	**2.79**	n.a.	n.a.	0.39

Growth Causes Energy

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	1.45	−1.57	**2.53**	**3.51**	1.15	**2.67**	−1.58	0.26
Capital	0.62	−2.05	−0.67	1.40	0.44	0.44	n.a.	0.90
Price	**3.58**	0.62	**3.83**	**3.78**	**3.65**	n.a.	**2.00**	**2.91**
Time trend	1.02	−0.35	**2.18**	**2.21**	0.62	**3.19**	−0.01	−0.70
Employment	−0.73	−1.79	**2.41**	n.a.	−0.70	n.a.	n.a.	−2.51

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

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 is presumably due to omitted variables bias. Perhaps because the elasticity of substitution between energy and capital is low it is hard to find an effect of energy on output when we control for capital. Similarly, if only changes in energy use due to changes in energy prices cause GDP, controlling for energy prices will remove the effect. Adding time trends or employment however, increases the significance of the "genuine" effects.

In the growth causes energy direction the significant effects are also no longer present when capital is added to the model. 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. For the Granger and Toda-Yamamoto procedures there is only a genuine effect in the growth causes energy direction when energy prices are controlled for. Also, any causality that is found in the cointegrated models is long run and there is no information in the short run parameters alone. We also carried out these tests for OECD and non-OECD countries separately (reported in the Appendix). Again, the results do not differ much from those reported here.

7. 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 to test whether genuine statistically significant effects exist in this literature.

We find a genuine effect from output to energy use in models that include energy prices as a control variable. A genuine causal effect also seems apparent from energy to output when employment is controlled for. This effect is more ambiguous because it is only present when cointegration test screening is used and cointegration is found whereas we find an effect from growth to energy when energy prices are controlled for 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. Controlling for capital seems to reduce the significance of the energy causes output tests.

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.[7] However the significance level for this test was only 5.5% in a one-tailed test.

Of course, we may fail to detect other genuine effects in the literature because of shortcomings in the underlying studies that we cannot correct in the meta-analysis. These may include: insufficiently frequent observations (Granger, 1988), errors in measurement, including measuring energy using heat equivalents rather than quality-adjusted measures (Stern and Enflo, 2013), nonlinearity of the energy-output relationship (Stern and Enflo, 2013; Sugihara et al., 2012), and other model misspecifications that go beyond omitted-variables bias.

Future research should include more individual studies with very long time series—there are currently only two studies (Vaona, 2012; Stern and Enflo, 2013) with time series with more than one hundred observations as well as studies with higher frequency data. Quality adjusted energy use should be a better measure of energy as a factor of production (Stern, 2011) but we were unable to test its effects due to a paucity of studies. More studies using quality adjusted energy use would, therefore, be helpful. More thought also needs to go into what variables should be controlled for and how to model the energy-output relationship. Meta-analysis should also be applied to subsets of the data with more consistently defined variables as well as to studies using panel data, which were deliberately excluded from the current analysis.

The meta-Granger causality tests used in this paper can also be applied in other research literatures where Granger causality testing is common. We have some general recommendations for such future studies. We recommend to convert all Granger causality test statistics to normal variates and to control for 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 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. 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. The slope-coefficient in the weighted least squares version of FAT-PET measures the genuine effect by exploiting the power trace while the constant is a test of publication and misspecification bias.

ACKNOWLEDGMENTS

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 Mo-

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

neta, Brandon Tracey, Tracy Wang, participants at the MAER-Net Symposium in Perth in September 2012, and three anonymous referees 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".

REFERENCES

Abosedra, S. and H. Baghestani, (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-Sahara African Countries." *Energy Economics* 30(5): 2391–2400. http://dx.doi.org/10.1016/j.eneco.2008.01.008.

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. http://dx.doi.org/10.1016/j.jpolmod.2009.03.004.

Alam, M., I. Begum, J. Buysse, S. Rahman, and G. Van Huylenbroeck (2011). "Dynamic Modeling of Causal Relationship between Energy Consumption, CO_2 Emissions and Economic Growth in India." *Renewable and Sustainable Energy Reviews* 15(6): 3243–3251. http://dx.doi.org/10.1016/j.rser.2011.04.029.

Altinay, G. and E. Karagol (2005). "Electricity Consumption and Economic Growth: Evidence from Turkey." *Energy Economics* 27(6): 849–856. http://dx.doi.org/10.1016/j.eneco.2005.07.002.

Ang, J. B. (2008a). "A Survey of Recent Developments in the Literature of Finance and Growth." *Journal of Economic Surveys* 22(3): 536–576. http://dx.doi.org/10.1111/j.1467-6419.2007.00542.x.

Ang, J. B. (2008b). "Economic Development, Pollutant Emissions and Energy Consumption in Malaysia." *Journal of Policy Modeling* 30(2): 271–278. http://dx.doi.org/10.1016/j.jpolmod.2007.04.010.

Azomahou, T., F. Laisney, and P. Nguyen Van (2006). "Economic Development and CO_2 Emissions: A Nonparametric Panel Approach." *Journal of Public Economics* 90(6–7): 1347–1363. http://dx.doi.org/10.1016/j.jpubeco.2005.09.005.

Belloumi, M. (2009). "Energy Consumption and GDP in Tunisia: Cointegration and Causality Analysis." *Energy Policy* 37(7): 2745–2753. http://dx.doi.org/10.1016/j.enpol.2009.03.027.

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. http://dx.doi.org/10.1109/EEM.2008.4579055.

Bowden, N. and J. Payne (2009). "The Causal Relationship between US Energy Consumption and Real Output: A Disaggregated Analysis." *Journal of Policy Modeling* 31(2): 180–188. http://dx.doi.org/10.1016/j.jpolmod.2008.09.001.

Bressler, S. L. and A. K. Seth (2011). "Wiener-Granger Causality: A Well Established Methodology." *NeuroImage* 58(2): 323–329. http://dx.doi.org/10.1016/j.neuroimage.2010.02.059.

Bruns, S. B. and C. Gross (2013). "What If Energy Time Series Are Not Independent? Implications For Energy-GDP Causality Analysis." *Energy Economics*.

Card, D. and A. B. Krueger (1995). "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* 85(2): 238–243. http://dx.doi.org/10.1016/j.eneco.2013.08.020.

Chang, C. and C. Soruco Carballo (2011). "Energy Conservation and Sustainable Economic Growth: The Case of Latin America and the Caribbean." *Energy Policy* 39(7): 4215–4221. http://dx.doi.org/10.1016/j.enpol.2011.04.035.

Chebbi, H. (2009). "Investigating Linkages between Economic Growth, Energy Consumption and Pollutant Emissions in Tunisia." International Association of Agricultural Economists Conference, Beijing, August 16–22.

Chen, P.-Y., S.-T. Chen, and C.-C. Chen (2012). "Energy Consumption and Economic Growth—New Evidence from Meta Analysis." *Energy Policy* 44: 245–255. http://dx.doi.org/10.1016/j.enpol.2012.01.048.

Chiou-Wei, S., C. Chen, and Z. Zhu (2008). "Economic Growth and Energy Consumption Revisited— Evidence from Linear and Nonlinear Granger Causality." *Energy Economics* 30(6): 3063–3076. http://dx.doi.org/10.1016/j.eneco.2008.02.002.

Chontanawat, J., L. Hunt, L., and R. Pierse (2008). "Does Energy Consumption Cause Economic Growth? Evidence from a Systematic Study of Over 100 Countries." *Journal of Policy Modeling* 30(2): 209–220. http://dx.doi.org/10.1016/j.jpolmod.2006.10.003.

Ciarreta, A., J. Otaduy, and A. Zarraga (2009). "Causal Relationship between Electricity Consumption and GDP in Portugal: A Multivariate Approach." *The Empirical Economics Letters* 8(7): 693–701.

Clarke, J. A. and S. Mirza (2006). "A Comparison of Some Common Methods for Detecting Granger Noncausality." *Journal of Statistical Computation and Simulation* 76: 207–231. http://dx.doi.org/10.1080/10629360500107741.

Engle, R. E. and C. W. J. Granger,(1987). "Cointegration and Error-Correction: Representation, Estimation, and Testing." *Econometrica* 55: 251–276. http://dx.doi.org/10.2307/1913236.

Erol, U., Yu, E. (1987). "Time Series Analysis of the Causal Relationships between US Energy and Employment." *Resources and Energy* 9(1): 75–89. http://dx.doi.org/10.1016/0165-0572(87)90024-7.

Esso, L. J. (2010). "Threshold Cointegration and Causality Relationship between Energy Use and Growth in Seven African Countries." *Energy Economics* 32(6): 1383–1391. http://dx.doi.org/10.1016/j.eneco.2010.08.003.

Fallahi, F. (2011). "Causal Relationship between Energy Consumption (EC) and GDP: A Markov-Switching (MS) Causality." *Energy* 36(7): 4165–4170. http://dx.doi.org/10.1016/j.energy.2011.04.027.

Ghosh, S. (2002). "Electricity Consumption and Economic Growth in India." *Energy Policy* 30(2): 125–129. http://dx.doi.org/10.1016/S0301-4215(01)00078-7.

Glasure, Y. U. (2002). "Energy and National Income in Korea: Further Evidence on the Role of Omitted Variables." *Energy Economics* 24(4): 355–365. http://dx.doi.org/10.1016/S0140-9883(02)00036-1.

Glasure, Y. U. and A.-R. Lee (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. http://dx.doi.org/10.1016/S0928-7655(96)00016-4.

Golam Ahamad, M. and A. Nazrul Islam (2011). "Electricity Consumption and Economic Growth Nexus in Bangladesh: Revisited Evidences." *Energy Policy* 39(10): 6145–6150. http://dx.doi.org/10.1016/j.enpol.2011.07.014.

Gonzalo, J. and J.-Y. Pitarakis (2002). "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. http://dx.doi.org/10.2307/1912791.

Granger, C. W. J. (1988). "Some Recent Developments in a Concept of Causality." *Journal of Econometrics* 39: 199–211. http://dx.doi.org/10.1016/0304-4076(88)90045-0.

Granger, C. W. J. and P. Newbold (1974). "Spurious Regressions in Econometrics." *Journal of Econometrics* 2: 111–120. http://dx.doi.org/10.1016/0304-4076(74)90034-7.

Gross, C. (2012). "Explaining the (Non-) Causality between Energy and Economic growth in the U.S.—A Multivariate Sectoral Analysis." *Energy Economics* 34(2): 489–499. http://dx.doi.org/10.1016/j.eneco.2011.12.002.

Hacker, R. S. and A. Hatemi-J (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. http://dx.doi.org/10.1080/02664760801920473.

Hamilton, J. D. (2009). "Causes and Consequences of the Oil Shock of 2007–08." *Brookings Papers on Economic Activity* 2009(1): 215–261. http://dx.doi.org/10.3386/w15002.

Hondroyiannis, G., S. Lolos, and E. Papapetrou (2002). "Energy Consumption and Economic Growth: Assessing the Evidence from Greece." *Energy Economics* 24(4): 319–336. http://dx.doi.org/10.1016/S0140-9883(02)00006-3.

Hsiao, C. (1979). "Autoregressive Modeling of Canadian Money and Income Data." *Journal of the American Statistical Association* 74: 553–560. http://dx.doi.org/10.1080/01621459.1979.10481651.

Ioannidis, J. P. A. (2005). "Why Most Published Research Findings Are False." *PLoS Medicine* 2(8): e124. http://dx.doi.org/10.1371/journal.pmed.0020124.

Jamil, F. and E. Ahmad (2010). "The Relationship between Electricity Consumption, Electricity Prices and GDP in Pakistan." *Energy Policy* 38(10): 6016–6025. http://dx.doi.org/10.1016/j.enpol.2010.05.057.

Jamil, F. and E. Ahmad (2011). "Income and Price Elasticities of Electricity Demand: Aggregate and Sector-Wise Analyses." *Energy Policy* 39(9): 5519–5527.

Jobert, T. and F. Karanfil (2007). "Sectoral Energy Consumption by Source and Economic Growth in Turkey." *Energy Policy* 35(11): 5447–5456. http://dx.doi.org/10.1016/j.enpol.2007.05.008.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12(2–3): 231–254. http://dx.doi.org/10.1016/0165-1889(88)90041-3.

Johansen, S. (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59(6): 1551–1580. http://dx.doi.org/10.2307/2938278.

Jumbe, C. (2004). "Cointegration and Causality between Electricity Consumption and GDP: Empirical Evidence from Malawi." *Energy Economics* 26(1): 61–68. http://dx.doi.org/10.1016/S0140-9883(03)00058-6.

Kaplan, M., I. Ozturk, and H. Kalyoncu (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. http://dx.doi.org/10.1016/j.enpol.2008.04.002.

Kaufmann, R. K. and D. I. Stern (1997). "Evidence for Human Influence on Climate from Hemispheric Temperature Relations." *Nature* 388: 39–44. http://dx.doi.org/10.1038/40332.

Lee, C. (2006). "The Causality Relationship between Energy Consumption and GDP in G-11 Countries Revisited." *Energy Policy* 34(9): 1086–1093. http://dx.doi.org/10.1016/j.enpol.2005.04.023.

Lee, T.-H. and W. Yang (2012). "Money-Income Granger-Causality in Quantiles." *Advances in Econometrics* 30: 385–409. http://dx.doi.org/10.1108/S0731-9053(2012)0000030017.

Lorde, T., K. Waithe, and B. Francis (2010). "The Importance of Electrical Energy for Economic Growth in Barbados." *Energy Economics* 32(6): 1411–1420. http://dx.doi.org/10.1016/j.eneco.2010.05.011.

Lotfalipour, M., M. Falahi, and M. Ashena (2010). "Economic Growth, CO_2 Emissions, and Fossil Fuels Consumption in Iran." *Energy* 35(12): 5115–5120. http://dx.doi.org/10.1016/j.energy.2010.08.004.

Masih, A. M. M. and R. Masih (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. http://dx.doi.org/10.1016/0140-9883(96)00009-6.

Masih, A. M. M. and R. Masih (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. http://dx.doi.org/10.1080/000368498324904.

Mehrara, M. (2007). "Energy Consumption and Economic Growth: The Case of Oil Exporting Countries." *Energy Policy* 35(5): 2939–2945. http://dx.doi.org/10.1016/j.enpol.2006.10.018.

Menyah, K. and Y. Wolde-Rufael (2010a). "CO_2 Emissions, Nuclear Energy, Renewable Energy and Economic Growth in the US." *Energy Policy* 38(6): 2911–2915. http://dx.doi.org/10.1016/j.enpol.2010.01.024.

Menyah, K. and Y. Wolde-Rufael (2010b). "Energy Consumption, Pollutant Emissions and Economic Growth in South Africa." *Energy Economics* 32(6): 1374–1382. http://dx.doi.org/10.1016/j.eneco.2010.08.002.

Mozumder, P. and A. Marathe (2007). "Causality Relationship between Electricity Consumption and GDP in Bangladesh." *Energy Policy* 35(1): 395–402. http://dx.doi.org/10.1016/j.enpol.2005.11.033.

Nickelsburg, G. (1985). "Small-Sample Properties of Dimensionalitv Statistics for Fitting VAR Models to Aggregate Economic Data." *Journal of Econometrics* 28: 183–192. http://dx.doi.org/10.1016/0304-4076(85)90118-6.

Oh, W. and K. Lee (2004). "Causal Relationship between Energy Consumption and GDP Revisited: The Case of Korea 1970–1999." *Energy Economics* 26(1): 51–59. http://dx.doi.org/10.1016/S0140-9883(03)00030-6.

Ozturk, I. (2010). "A Literature Survey on Energy-Growth Nexus." *Energy Policy* 38(1): 340–349. http://dx.doi.org/10.1016/j.enpol.2009.09.024.

Pao, H. and C. Tsai (2011). "Modeling and Forecasting the CO_2 Emissions, Energy Consumption, and Economic Growth in Brazil." *Energy* 36(5): 2450–2458. http://dx.doi.org/10.1016/j.energy.2011.01.032.

Paul, B. and G. Uddin (2010). "Energy and Output Dynamics in Bangladesh." *Energy Economics* 33(3): 480–487. http://dx.doi.org/10.1016/j.eneco.2010.11.011.

Paul, S. and R. N. Bhattacharya (2004). "Causality between Energy Consumption and Economic Growth in India: A Note on Conflicting Results." *Energy Economics* 26(6): 977–983. http://dx.doi.org/10.1016/j.eneco.2004.07.002.

Payne, J. E. (2009). "On the Dynamics of Energy Consumption and Output in the US." *Applied Energy* 86(4): 575–577. http://dx.doi.org/10.1016/j.apenergy.2008.07.003.

Payne, J. E. (2010a). "Survey of the International Evidence on the Causal Relationship between Energy Consumption and Growth." *Journal of Economic Studies* 37(1): 53–95. http://dx.doi.org/10.1108/01443581011012261.

Payne, J. E. (2010b). "A Survey of the Electricity Consumption-Growth Literature." *Applied Energy* 87(3): 723–731. http://dx.doi.org/10.1016/j.apenergy.2009.06.034.

Pesaran, M. H. and Y. Shin (1999). "An Autoregressive Distributed Lag Modelling Approach to Cointegrated Analysis." In S. Strom, ed, *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Cambridge: Cambridge University Press.

Pesaran, M. H., Y. Shin, and R. J. Smith (2001). "Bounds Testing Approaches to the Analysis of Level Relationships." *Journal of Applied Econometrics* 16(3): 289–326. http://dx.doi.org/10.1002/jae.616.

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. and R. Salim (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. http://dx.doi.org/10.1108/17468801111104377.

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. http://dx.doi.org/10.2307/2078071.

Sa'ad, S. (2010). "Energy Consumption and Economic Growth: Causality Relationship for Nigeria." *OPEC Energy Review* 34(1): 15–24. http://dx.doi.org/10.1111/j.1753-0237.2010.00173.x.

Salim, R., S. Rafiq, and A. Hassan (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. and U. Soytas (2009). "Are Global Warming and Economic Growth Compatible? Evidence from Five OPEC Countries." *Applied Energy* 86(10): 1887–1893. http://dx.doi.org/10.1016/j.apenergy.2008.12.007.

Shiu, A. and P.-L. Lam (2004). "Electricity Consumption and Economic Growth in China." E*nergy Policy* 32(1): 47–54. http://dx.doi.org/10.1016/S0301-4215(02)00250-1.

Sims, C. (1972). "Money, Income, and Causality." *American Economic Review* 62(4): 540—552.

Soytas, U. and R. Sari (2009). "Energy Consumption, Economic Growth, and Carbon Emissions: Challenges Faced by an EU Candidate Member." *Ecological Economics* 68(6): 1667–1675. http://dx.doi.org/10.1016/j.ecolecon.2007.06.014.

Soytas, U., R. Sari, and B. Ewing (2007). "Energy consumption, income, and carbon emissions in the United States." *Ecological Economics* 62(3–4): 482–489. http://dx.doi.org/10.1016/j.ecolecon.2006.07.009.

Stanley, T. D. (2005a). "Beyond Publication Bias." *Journal of Economic Surveys* 19(3): 309–345. http://dx.doi.org/10.1111/j.0950-0804.2005.00250.x.

Stanley, T. D. (2005b). "Integrating the Empirical Tests of the Natural Rate Hypothesis: A Meta-Regression Analysis." *Kyklos* 58(4): 611–634. http://dx.doi.org/10.1111/j.0023-5962.2005.00305.x.

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. and H. Doucouliagos (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. and H. Doucouliagos (2012). *Meta-Regression Analysis in Economics and Business*. London: Routledge.

Stern, D. I. (1993). "Energy Use and Economic Growth in the USA: A Multivariate Approach." *Energy Economics* 15: 137–150. http://dx.doi.org/10.1016/0140-9883(93)90033-N..

Stern, D. I. (2000). "A Multivariate Cointegration Analysis of the Role of Energy in the U.S. Macroeconomy." *Energy Economics* 22: 267–283. http://dx.doi.org/10.1016/S0140-9883(99)00028-6.

Stern, D. I. (2011). "The Role of Energy in Economic Growth." *Annals of the New York Academy of Sciences* 1219: 26–51. http://dx.doi.org/10.1111/j.1749-6632.2010.05921.x.

Stern, D. I. and K. Enflo (2013). "Causality between Energy and Output in the Long-Run." *Energy Economics* 39: 135–146. http://dx.doi.org/10.1016/j.eneco.2013.05.007.

Sugihara, G., R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch (2012). "Detecting Causality in Complex Ecosystems." *Science* 338: 496–500. http://dx.doi.org/10.1126/science.1227079.

Toda, H. Y. and T. Yamamoto (1995). "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes." *Journal of Econometrics* 66(1–2): 225–250. http://dx.doi.org/10.1016/0304-4076(94)01616-8.

Türkekul, B. and G. Unakıtan (2011). "A Co-Integration Analysis of the Price and Income Elasticities of Energy Demand in Turkish Agriculture." *Energy Policy* 39(5): 2416–2423. http://dx.doi.org/10.1016/j.enpol.2011.01.064.

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. http://dx.doi.org/10.1016/j.enpol.2012.02.023.

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. http://dx.doi.org/10.1080/15567249.2011.563261.

Vlahinic-Dizdarevic, N. and S. Zikovic (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. http://dx.doi.org/10.1016/j.eneco.2008.11.005.

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. http://dx.doi.org/10.1016/j.enpol.2009.08.053.

Wolde-Rufael, Y. (2010b). "Coal Consumption and Economic Growth Revisited." *Applied Energy* 87(1): 160–167. http://dx.doi.org/10.1016/j.apenergy.2009.05.001.

Wolde-Rufael, Y., K. Menyah (2010). "Nuclear Energy Consumption and Economic Growth in Nine Developed Countries." *Energy Economics* 32(3): 550–556. http://dx.doi.org/10.1016/j.eneco.2010.01.004.

Yoo, S.-H. (2006). "The Causal Relationship between Electricity Consumption and Economic Growth in the ASEAN Countries." *Energy Policy* 34(18): 3573–3582. http://dx.doi.org/10.1016/j.enpol.2005.07.011.

Yoo, S. and S. Ku (2009). "Causal Relationship between Nuclear Energy Consumption and Economic Growth: A Multi-Country Analysis." *Energy Policy* 37(5): 1905–1913. http://dx.doi.org/10.1016/j.enpol.2009.01.012.

Yoo, S. and S. Kwak (2010). "Electricity Consumption and Economic Growth in Seven South American Countries." *Energy Policy* 38(1): 181–188. http://dx.doi.org/10.1016/j.enpol.2009.09.003.

Yu, E. S. H. and B.-K. Hwang (1984). "The Relationship between Energy and GNP: Further Results." *Energy Economics* 6(3): 186–190. http://dx.doi.org/10.1016/0140-9883(84)90015-X.

Yu, E. and J.-Y. Choi (1985). "The Causal Relationship between Energy and GNP: An International Comparison." *The Journal of Energy and Development* 10(2): 249–272.

Yuan, J., C. Zhao, S. Yu, and Z. Hu (2007). "Electricity Consumption and Economic Growth in China: Cointegration and Co-Feature Analysis." *Energy Economics* 29(6): 1179–1191. http://dx.doi.org/10.1016/j.eneco.2006.09.005.

Yusof, N. and N. Latif (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. http://dx.doi.org/10.1016/j.eneco.2007.05.001.

Zachariadis, T. and N. Pashourtidou (2007). "An Empirical Analysis of Electricity Consumption in Cyprus." *Energy Economics* 29(2): 183–198. http://dx.doi.org/10.1016/j.eneco.2006.05.002.

Zamani, M. (2007). "Energy Consumption and Economic Activities in Iran." *Energy Economics* 29(6): 1135–1140. http://dx.doi.org/10.1016/j.eneco.2006.04.008.

Zapata, H. O. and A. N. Rambaldi (1997). "Monte Carlo Evidence on Cointegration and Causation." *Oxford Bulletin of Economics and Statistics* 59(2): 285–298. http://dx.doi.org/10.1111/1468-0084.00065.

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. and X.-M. Cheng (2009). "Energy Consumption, Carbon Emissions, and Economic Growth in China." *Ecological Economics* 68(10): 2706–2712. http://dx.doi.org/10.1016/j.ecolecon.2009.05.011.

Zhao, Z. and J. Yuan (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. http://dx.doi.org/10.1016/j.enpol.2009.01.048.

Zou, G. and K. Chau (2006). "Short-and Long-Run Effects Between Oil Consumption and Economic Growth in China." *Energy Policy* 34(18): 3644–3655. http://dx.doi.org/10.1016/j.enpol.2005.08.009.

APPENDIX 1: DIFFERENCES BETWEEN OECD AND NON-OECD COUNTRIES

In this Appendix we present separate versions of Tables 7 and 8 for OECD and non-OECD countries. Comparing Tables A7a and A7b with Table 7 in the main body of the paper, the results for the two groups of countries are mostly similar. The most noticeable differences here are that for the OECD countries: there are no genuine effects for the long run causality tests using cointegration models; there is a genuine effect from energy to output when a time trend is included; and there is a genuine effect from energy to growth when price is controlled for and only a weak effect in the opposite direction. For the non-OECD countries, the most interesting difference is that energy causes output when employment is controlled for. But each of these results depends mainly on the results of a single study (in the case of long run causality, Zachariadis, 2007) or a small number of observations and hence the reliability of these differences is limited.

Looking at Tables A8a and A8b, the main difference between the results reported here and in the main paper is that for the OECD countries there is a very significant genuine effect in the energy causes growth direction when we control for prices, though this is based on a small number of observations. Also for the joint test there is a significant effect when capital is controlled for. But this result is due to a single very significant t-test from Kaplan et al. (2011) for Turkey, which was a member of the OECD in 1990. For the non-OECD countries there are mostly significant effects in the energy causes growth direction when controlling for employment and in the growth causes energy direction when controlling for prices.

Table A7a: OECD 1990 Countries

	Techniques								
Methodology	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto	Hsiao
Energy Causes Growth									
Levels coefficient	−0.158	−1.200	4.162	−0.962	−0.158	n.a.	0.348	1.070	3.398
	(−0.096)	(−0.759)	(1.872)	(−0.269)	(−0.096)		(0.472)	(1.655)	(1.516)
Joint slope coefficient	0.376	0.400	−0.322	0.586	0.376	n.a.	−0.005	−0.095	−0.487
	(1.372)	(1.584)	(−0.910)	(0.992)	(1.372)		(−0.036)	(−0.961)	(−1.266)
Growth Causes Energy									
Levels coefficient	−0.433	1.743	3.662	−5.608	−0.433	n.a.	1.858	1.386	2.181
	(−0.249)	(1.108)	(1.048)	(−1.607)	(−0.249)		(3.597)	(1.203)	(1.191)
Joint slope coefficient	0.445	−0.196	**−0.138**	1.450	0.445	n.a.	−0.309	−0.096	−0.187
	(1.603)	(−0.763)	**(−0.240)**	(2.551)	(1.603)		(−3.106)	(−0.484)	(−0.568)

	Variables				Variable Definition	
Methodology	Time	Price	Capital	Employment	Macro-macro	MM Total Energy
Energy Causes Growth						
Levels coefficient	−0.399	−9.631	3.927	3.927	1.214	0.229
	(−0.280)	(−2.549)	(3.226)	(3.226)	(2.631)	(0.249)
Joint slope coefficient	**0.504**	**2.370**	−0.470	−0.470	−0.046	0.119
	(1.941)	**(3.217)**	(−1.959)	(−1.959)	(−0.641)	(0.775)
Growth Causes Energy						
Levels coefficient	7.639	−6.312	4.470	4.470	1.142	0.340
	(2.199)	(−0.714)	(2.925)	(2.925)	(2.207)	(0.368)
Joint slope coefficient	−1.288	**1.584**	−0.583	−0.583	0.017	0.172
	(−2.036)	**(0.928)**	(−1.991)	(−1.991)	(0.216)	(1.150)

Notes: t-stats in parentheses.
Notable differences to Table 7 in the main paper are marked in bold.

Table A7b: Non-OECD Countries

	Techniques								
Methodology	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto	Hsiao
Energy Causes Growth									
Levels coefficient	1.411	0.998	1.653	1.944	1.308	1.372	−0.105	0.741	1.293
	(2.967)	(1.852)	(1.795)	(1.396)	(2.514)	(2.768)	(−0.068)	(0.540)	(1.160)
Joint slope coefficient	−0.001	0.034	0.000	−0.131	0.011	0.084	0.194	−0.026	−0.120
	(−0.010)	(0.255)	(0.002)	(−0.497)	(0.101)	(0.733)	(0.552)	(−0.094)	(−0.467)
Growth Causes Energy									
Levels coefficient	0.522	2.147	−0.156	−2.507	0.699	−1.780	−0.274	1.589	2.682
	(0.550)	(2.046)	(−0.169)	(−1.110)	(0.663)	(−1.873)	(−0.240)	(0.939)	(1.645)
Joint slope coefficient	0.154	−0.306	0.447	0.650	0.123	0.669	0.253	−0.135	−0.402
	(0.847)	(−1.558)	(2.523)	(1.420)	(0.624)	(3.60)	(0.895)	(−0.385)	(−1.136)

	Variables				Variable Definition	
Methodology	Time	Price	Capital	Employment	Macro-macro	MM Total Energy
Energy Causes Growth						
Levels coefficient	1.354	1.057	0.420	−1.331	0.589	0.457
	(1.158)	(1.224)	(0.280)	(−1.030)	(0.631)	(0.378)
Joint slope coefficient	−0.054	−0.000	0.200	**0.577**	0.089	0.095
	(−0.202)	(−0.000)	(0.627)	**(2.328)**	(0.442)	(0.358)
Growth Causes Energy						
Levels coefficient	1.718	−1.568	1.825	1.732	0.437	1.556
	(1.248)	(−1.435)	(1.585)	(1.219)	(0.647)	(1.963)
Joint slope coefficient	−0.069	0.498	−0.081	−0.166	0.117	−0.133
	(−0.212)	(2.623)	(−0.410)	(−0.595)	(0.777)	(−0.765)

Notes: t-stats in parentheses.
Notable differences to Table 7 in the main paper are marked in bold.

Table A8a: OECD Countries

Energy Causes Growth

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	1.454	0.921	0.430	**5.576**	1.454	n.a.	−0.715	−0.502
Capital	0.582	0.769	0.521	**2.391**	0.582	n.a.	n.a.	−0.988
Price	**1.654**	0.446	**4.417**	**6.493**	**1.654**	n.a.	n.a.	1.574
Time trend	**2.005**	0.882	**2.184**	**6.970**	**2.005**	n.a.	0.862	0.975
Employment	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Growth Causes Energy

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	0.926	−1.129	0.921	**2.390**	0.926	n.a.	−3.875	0.492
Capital	−0.159	−2.897	0.769	0.220	−0.159	n.a.	n.a.	−1.536
Price	**1.659**	1.171	0.446	1.371	**1.659**	n.a.	n.a.	**1.933**
Time trend	0.550	−1.263	0.882	−0.818	0.550	n.a.	−1.308	−3.123
Employment	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

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

Table A8b: Non-OECD Countries

Energy Causes Growth

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	0.773	0.222	1.309	0.701	0.693	0.498	0.996	0.363
Capital	−1.507	−1.341	−2.664	−0.565	−0.857	n.a.	n.a.	0.466
Price	0.789	−0.345	−0.005	0.107	0.713	n.a.	**1.909**	n.a.
Time trend	0.828	−1.158	0.836	n.a.	n.a.	0.739	0.182	−0.279
Employment	**2.366**	**2.250**	**3.899**	n.a.	**2.362**	n.a.	n.a.	1.060

Growth Causes Energy

Controls	Cointegration	Short run	Long run	Joint	Johansen	Engle-Granger	Granger	Toda-Yamamoto
None	0.232	−1.431	1.586	0.460	−0.038	**2.416**	0.547	−0.138
Capital	0.864	−1.380	−0.758	0.559	0.578	0.949	n.a.	0.499
Price	**2.677**	0.370	**3.522**	**1.867**	**2.858**	n.a.	**2.572**	n.a.
Time trend	0.698	−1.213	**3.431**	n.a.	n.a.	**2.741**	1.063	0.814
Employment	−2.808	−1.123	0.693	n.a.	−3.385	n.a.	n.a.	−3.178

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

REFERENCES

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.

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. http://dx.doi.org/10.1016/j.eneco.2007.05.001.