CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

The case of coal: A meta analysis of demand and substitution

Bachelor's thesis

Author: Lucia Sihelská
Study program: Ekonomická teorie (B6201)
Supervisor: doc. PhDr. Zuzana Havránková, Ph.D.
Year of defense: 2022

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 2, 2022

Lucia Sihelská

Abstract

Interfuel elasticities capturing energy substitution possibilities are key parameters used to assess the costs of various environmental policies trying to mitigate climate change. In this thesis, special attention is dedicated to coal, the most polluting energy input. We collect and synthesize 893 estimates of the own-price elasticity of demand for coal and cross-price elasticity of demand related to coal and three alternative fuels – electricity, gas, and oil. The data are collected from 43 studies. State-of-the-art meta-analytic tools are employed to explore what drives the estimates of the elasticities. To detect publication bias, we use both linear and non-linear methods. We find that publication bias affects the estimates of the own-price elasticity for coal and cross-price elasticity between three alternative fuels and coal, which causes the reported evidence to be significantly exaggerated. In contrast, substantial publication bias is not detected in the case of cross-price elasticity between coal and alternative fuels. Based on this, we demonstrate that the substitution of other fuels for coal is feasible, although limited. To explain heterogeneity in the estimates, we use Bayesian model averaging. Other factors that systematically affect the estimated elasticities include the estimation techniques used to generate the estimates and whether the author controlled for potential sources of endogeneity.

JEL Classification	C83, Q40, Q41
Keywords	meta-analysis, publication bias, coal, elasticity, substitution
Title	The case of coal: A meta analysis of demand and substitution

Abstrakt

Elasticita medzi rôznymi druhmi palív patrí medzi kľúčové parametre, ktoré sa používajú pri stanovení nákladov enviromentálnych politík snažiacich sa o boj s klimatickou krízou. Táto práca venuje osobitú pozornosť uhliu, najviac znečisťujúcemu druhu paliva. Zozbierali a syntetizovali sme 893 odhadov cenovej elasticity dopytu po uhlí a krížovej elasticity dopytu medzi uhlím a troma alternatívnymi palivami – elektrinou, zemným plynom a ropou. Pozorovania sú zhromaždené zo 43 štúdií. Na základe najmodernejších meta-analytických metód skúmame premenné ovplyvňujúce odhady estlaticity. Na odhalenie publikačnej selektivity používame lineárne aj nelineárne techniky. Publikačná selektivita je prítomná v prípade cenovej elasticity dopytu po uhlí a krížovej elasticity dopytu medzi alternatívnymi palivami a uhlím, v dôsledku ktorej sú prezentované odhady značne zveličené. Naopak, výrazná publikačná selektivita nebola detekovaná v prípade krížovej elasticity dopytu medzi uhlím a alternatívnymi palivami. Na základne toho demonštrujeme, že substitúcia uhlia inými palivami je možná, hoci obmedzená. Za účelom vysvetlenia heterogenity v odhadoch používame Bayesovské priemerovanie modelov. Medzi ďalšie faktory systematicky ovplyvňujúce odhady elasticity patria techniky odhadu a to, či autor kotroloval potenciálne zdroje endogenity.

Klasifikácia JEL	C83, Q40, Q41
Kľúčové slová	meta-analýza, publikačná selektivita, uhlie, elasticita, substitúcia
Názov práce	Případ uhlí: Meta-analýza poptávky a substituce

Acknowledgments

I would like to express my greatest gratitude to doc. PhDr. Zuzana Havránková, Ph.D. for introducing me to the topic of meta-analysis, for her guidance and detailed comments throughout the whole semester. I would also like to thank prof. David Stern for kindly providing me with his dataset. Furthermore, I am grateful to Mgr. Dominika Ehrenbergerová for her valuable help.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Sihelská, Lucia: *The case of coal: A meta analysis of demand and substitution*. Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2022, pages 99. Advisor: doc. PhDr. Zuzana Havránková, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

BMA Bayesian Model Averaging

FAT Funnel Asymmetry Test

FE Fixed Effects

FMA Frequentist Model Averaging

GDP Gross Domestic Product

GMM Generalized Method of Moments

IV Instrumental Variable

ML Maximum Likelihood

OECD Organisation for Economic Co-operation and Development

PET Precision Effect Test

PIP Posterior Inclusion Probability

PM Posterior Mean

PMP Posterior Model Probabilities

PSD Posterior Standard Deviation

SD Standard Deviation

SE Standard Error

SUR Seemingly Unrelated Regression

UIP Unit Information Prior

WAAP Weighted Average of Adequately Powered

Chapter 1

Introduction

Coal consumption is a major source of carbon emissions (Hannah Ritchie and Rosado, 2020). Besides carbon's massive contribution to climate change, it is also linked to high social costs (Nordhaus, 2017). Moreover, coal combustion is also the cause of air pollution associated with respiratory, cardiovascular, and neurological diseases (Cohen et al., 2004). Thus, investigating possibilities to substitute for cleaner energy sources is of great importance in estimating the costs of mitigating the unwanted effects of coal consumption. In this work, we show that the demand for coal is less elastic than previously thought, as the exaggeration due to publication bias is at least two-fold. We also demonstrate that the substitution of cleaner fuels for coal is feasible, although limited.

In economics, the measure used to capture substitution possibilities is called elasticity. Various concepts of elasticity exist but its common purpose is to measure the change in the demand for a good in relation to price changes. Our focus therefore lies on the estimates of the fuel elasticities. These estimates are key parameters of many environmental models assessing the impact of various policies trying to tackle climate change. As demonstrated by Antimiani et al. (2015), such models show a certain degree of sensitivity to the fuel elasticities. The results produced by different sets of estimates vary significantly. As they argue, there is a great need for exact and accurate estimates of these important parameters to be able to determine the real costs of carbon reduction efforts.

The literature on the topic of interfuel substitution is rich. There are two narrative reviews of the literature, provided by Apostolakis (1990) and Bacon and Mundial (1992). They demonstrate a large variation in the estimates of interfuel elasticities. Indeed, the values are found to vary across countries (Griffin, 1977), industries (Uri, 1982c) and sectors (Shahiduzzaman and Alam, 2014). There are also two previous attempts to analyze why estimates differ quantitatively. The first can be found in Stern (2012), who explored the manufacturing-industry and country-level estimates of the symmetric interfuel elasticities. This paper was closely followed by Chen (2017), whose focus was on the fuel substitution in the electricity generation sector.

This work builds on the previous meta-analyses but dedicates special attention to coal, the largest pollutant of all fuels. Our main purpose is to present a systematic quantitative review of the literature dealing with the topic of coal substitution. We collect 893 estimates of the own-price elasticity of demand for coal and pairwise cross-price elasticities of substitution between coal and three alternative fuels – electricity, gas, and oil. The estimates are collected from 43 studies. We mentioned two existing meta-analyses on interfuel substitution, but to the best of our knowledge, we present the first meta-analysis of the own-price elasticity for coal. Our work is also the first attempt to quantitatively analyze estimates of the asymmetric cross-price elasticity instead of symmetric elasticity. This allows us to distinguish between the substitution of other fuels for coal and the substitution of coal for other fuels. We show that such distinction is essential as the two directions of the substitution differ. Next to the estimates, we collect their standard errors and other explanatory variables to investigate why the reported elasticities vary. We control for several important issues. Both aforementioned papers fail to use sophisticated, rigorous tools to account for publication bias. As noted: "left unaddressed, [publication] selectivity can lead to biased estimates and misleading confidence sets in published studies," (Andrews and Kasy, 2019, p. 1266). The problem of model uncertainty is not considered in these works at all, either. Thus, our main contribution is three-fold. Firstly, we conduct the first meta-analysis of the own-price elasticity of demand for coal. Secondly, we use rigorous, up-to-date methods to detect possible publication bias in the estimates of the own-price elasticity and cross-price elasticity of demand related to coal. Thirdly, we investigate why the estimates differ by employing model averaging techniques that treat for the model uncertainty.

We start by paying attention to potential publication bias. Publication bias, or selective reporting, is present if researchers systematically under-report estimates with certain characteristics. In such cases, the existing evidence becomes distorted and exaggerated. To test for publication bias, we use both linear and non-linear methods. We show that the own-price elasticity for coal and the cross-price elasticity between other fuels and coal is significantly affected by selective reporting. On the contrary, we find that the cross-price elasticity between coal and other fuels is not likely to suffer from substantial publication bias. Heterogeneity in the estimates is explored using Bayesian model averaging (BMA), which treats for the model uncertainty. Apart from selective reporting, the estimates are systematically impacted by the study settings and estimation techniques.

The thesis is structured as follows: Chapter 2 discusses several different concepts of the elasticity used to measure interfuel substitution as well as the estimation procedure and shows previous findings. Chapter 3 describes the data collection and presents the summary statistics of our dataset. In Chapter 4, we describe publication bias, use rigorous, state-of-the-art tools to detect its presence and discuss the results. Chapter 5 provides an in-depth analysis of the heterogeneity in the literature concerning coal substitution and a discussion of our findings. Chapter 6 presents the final concluding remarks. In Appendix A, Appendix B, and Appendix C, additional supporting materials are provided.

Chapter 2

Theoretical background

2.1 Defining the elasticity

The most common measures used in the literature concerning interfuel substitution are own and cross-price elasticity of demand η_{ij}, Allen-Uwaza elasticity of substitution σ_{ij} and Morishima elasticity of substitution σ_{ij}^{M}.

The own and cross-price elasticity of demand can be calculated as

$$\eta_{ii} = \frac{\delta ln X_i(y,p)}{\delta p_i}$$

$$\eta_{ij} = \frac{\delta ln X_i(y,p)}{\delta p_j} \quad for \quad i = j,$$

respectively, while keeping the output constant. X_i denotes the demand for input i, y is output, p is the vector of factor prices and p_i and p_j are corresponding prices for input i and input j, respectively (Stern, 2012).
The own-price elasticity measures the proportional change in demand for the input i when its price changes proportionally. Negative own-price elasticity of demand indicates that with increasing price, the demand for i decreases. Because such behaviour can be considered natural, it is often assumed that the own-price elasticity should lie in the interval (-∞, 0).
The cross-price elasticities are asymmetric so that $\eta_{ij} = \eta_{ji}$. Cross-price elasticity of demand measures the proportional change in demand for the input i when price for input j changes proportionally. A positive cross-price elasticity implies that with an increase in the price for input j, demand for input i increases. In this case, inputs i and j form substitutes. On the other hand,

a negative cross-price elasticity implies that increasing the price for input j results in lower demand for input i. In other words, a decrease in input j is accompanied by a decrease of input i, and these inputs are considered complements.

In addition to the elasticity of demand between different fuels, some researchers also estimate the elasticity of demand between energy, capital, labour, and potentially material. This allows them to derive the so-called total price elasticities for fuels, which can be computed as

$$\eta_{ii} = \eta_{ij} + \eta_{EE} S_i$$

$$\eta_{ij} = \eta_{ij} + \eta_{EE} S_j \quad for \quad i = j,$$

where η_{EE} denotes the own-price elasticity of demand for energy and S_i and S_j stand for cost shares of input i and j, respectively. This enables to account for the substitution between different fuels while the total quantity of energy consumed can vary. Total price elasticities are computed by, for example, Pindyck (1979).

Allen-Uwaza elasticities can be computed as

$$\sigma_{ii} = \frac{1}{S_i} \frac{\delta ln X_i(y,p)}{\delta p_i}$$

$$\sigma_{ij} = \frac{1}{S_j} \frac{\delta ln X_i(y,p)}{\delta p_j} \quad for \quad i = j,$$

while keeping the output constant. Allen-Uwaza elasticities are symmetric, so that $\sigma_{ij} = \sigma_{ji}$. Because the cost shares are non-negative, they have the same signs as the corresponding cross-price elasticities. We can therefore assume that the own-price Allen-Uwaza elasticity should also lie in the interval $(-\infty, 0)$. Similarly, if cross-price elasticity $\sigma_{ij} > 0$, x_i and x_j are substitutes, and if $\sigma_{ij} < 0$, x_i and x_j can be viewed as complements. Allen-Uwaza elasticity is reported by, for instance, Lin and Tian (2017) or Wang and Lin (2017).

However, Blackorby and Russell (1989) argue that the Allen-Uwaza elasticity of substitution becomes uninformative in case of more than two inputs, because it provides no additional information to that contained in the price elasticity of demand. Therefore, some studies, such as Kim (2019) or Shahiduzzaman and Alam (2014) estimate Morishima elasticity of substitution, which indicates whether the ratio of inputs increases or decreases when input prices change.

Morishima elasticity of substitution is defined as

$$\sigma_{ij}^{M} = \frac{\delta ln(X_j(y,p)/X_i(y,p))}{\delta ln(p_i/p_j)}|_{p_j} = \eta_{ji} - \eta_{ii} \quad for \quad i = j,$$

while keeping the output constant. Morishima elasticity of substitution is asymmetric, so that $\sigma_{ij}^{M} = \sigma_{ji}^{M}$. As noted, the Morishima elasticity of substitution σ_{ij}^{M} describes how the proportional change in the ratio of inputs reacts to the proportional change in the price for input i. Therefore, a positive Morishima elasticity of substitution means that the increase in the price for input i given price for input j declines the ratio $\frac{x_i}{x_j}$ and the inputs are said to be substitutes. Similarly, a negative Morishima elasticity of substitution implies that with increasing price for input i, the ratio $\frac{x_i}{x_j}$ increases too and the inputs can be considered complements (Kim, 2019).

Stern (2012) mentions yet another concept of elasticity, the so-called shadow elasticity, which is the average of Morishima elasticity. He argues that the shadow elasticities are good summary statistics of the overall substitutability among inputs, as they are symmetric, and thus, they are fewer in total numbers. He carries out a meta-analysis of shadow elasticities. This approach is followed by Chen (2017). Nonetheless, because most studies report own and cross-price elasticities of demand, we do not convert them into shadow elasticities. Firstly, the elasticity between coal and alternative fuels might differ both in sign and magnitude from the elasticity in the reversed direction. Thus, investigating symmetric elasticity might hide certain patterns and differences between the two directions of the elasticity, as this work demonstrates later on. Secondly, in order not to mix apples and oranges, we wanted to keep the estimates consistent. However, every recalculation of the elasticity into different metrics might introduce a certain degree of inaccuracy. Therefore, our further focus is only on the estimates of the own and cross-price elasticities of demand.

2.2 Estimating the elasticity

The most common approach prevailing in the reviewed literature is the translog framework, first introduced by Christensen et al. (1973).
As a second-order approximation of the arbitrary production function, it is not required to specify a particular production function. Moreover, the elasticity of substitution between inputs is allowed to vary (Lin and Tian, 2017). Most researchers prefer the translog cost function to the production function, for example, Griffin (1977), Harvey and Marshall (1991), Cho et al. (2004), or Shahiduzzaman and Alam (2014), among many others. The translog cost function can be derived as follows:

Assuming that the aggregate production function is weakly separable in inputs, using duality theory of cost and production functions and following Shephard (1953), the cost function is also assumed to be weakly separable and can be written as:

$$C_E = C(Y, P_{E_1}, ...P_{E_n})$$

where Y denotes the output and P_{E_i}, $i = 1...n$ is the price of the i^{th} fuel. Under the assumption that the energy input function is positive, monotonic and has curvature – the so-called regularity conditions, the energy cost function can be rewritten as:

$$C_E = Y \quad P_E(P_{E_1}, ...P_{E_n})$$

where P_E denotes the energy price aggregation function satisfying the regularity conditions (Diewert, 1973). P_E can then be represented by a translog unit cost function with constant returns to scale:

$$lnP_E = ln\alpha_0 + \sum_{i=1}^{n} \alpha_i lnP_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} lnP_i lnP_j \tag{2.1}$$

where α_0 denotes the state of technical knowledge and the remaining Greek letters are parameters to be estimated. The following parameter restrictions are imposed to ensure symmetry and homogeneity of degree one in prices :

$$\gamma_{ij} = \gamma_{ji}, \quad \sum_{i=1}^{n} \gamma_{ij} = 0, \quad \sum_{i=1}^{n} \gamma_i = 1$$

According to Shephard (1953), the partial derivative of the cost function with respect to the price of the given input yields conditional factor demands for inputs:

$$\frac{\delta P_E}{\delta P_{E_i}} = X_{E_i}$$

This is known as Shephard's lemma, which yields cost share equations when it becomes expressed in logarithmic form for the translog frontier:

$$\frac{\delta lnP_E}{\delta lnP_{E_i}} = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} lnP_j = S_i,$$

where S_i denotes the cost share of fuel i. The cost shares sum to unity (Uri, 1979a). Now, it is possible to calculate corresponding own and cross-price elasticities of demand as follows:

$$\eta_{ii} = \frac{\gamma_{ii} + S_i^2 - S_i}{S_i} \tag{2.2}$$

$$\eta_{ij} = \frac{\gamma_{ij} + S_i S_j}{S_i} \quad for \quad i = j \tag{2.3}$$

Though the translog framework prevails in the literature, a number of researchers choose a different approach. Jones (1995), Considine (2018), Urga and Walters (2003) and Steinbuks and Narayanan (2015) adopt the linear logit model. The differential fuel allocation model is used by Suh (2016). In addition, Pettersson et al. (2012) choose a generalized Leontief model to estimate the elasticity of demand between different energy inputs[1].

2.3 Previous findings

The estimates from the empirical studies differ greatly both in signs and magnitudes. Some researchers found fuel substitutability, while others discovered complementarity. The values vary significantly, too. In general, it is difficult to determine what drives the estimates and affects fuel substitution. A few studies trying to summarize the findings exist. The first narrative review of the topic is conducted by Apostolakis (1990), who includes 14 studies reporting the empirical estimates of the elasticity between coal, electricity, gas, and

[1] For the sake of simplicity, we will not further discuss these specifications. Detailed derivations of the models can be found in the aforementioned papers.

oil. The focus of this work is on the interfuel and capital-energy substitution in the manufacturing industry. In most cases, substitutability between fuels is detected. The second literature review can be found in Bacon and Mundial (1992), who include up to 10 studies and focus on the country-level estimates. They testify a large variation in the estimates of the elasticity across countries, but also over time. They also recognize that cross-sectional data are likely to yield larger estimates than time-series. Stern (2012) is the first to analyze this variation quantitatively. 369 estimates of the elasticity related to coal, electricity, gas, and oil from 47 studies are collected and recalculated into shadow elasticities for the purpose of this meta-analysis. He finds that the industrial-level estimates tend to be greater than the state-level estimates. In line with the previous literature review, he also finds that the estimates generated by cross-sectional data are the largest. The estimates produced by panel data are intermediate in magnitude and the estimates produced by time series attain the smallest values. This paper is closely followed by Chen (2017), who dedicates attention to the fuel substitution in the electricity generation sector. 32 estimates of the elasticity related to coal, gas, and oil were collected from 11 studies. Substitutability between the fuels is found to be greater in the US than in the remaining regions, including Europe, Mexico, Turkey, Japan, and Australia. However, few important features remain unaddressed by these meta-analyses. In this work, we focus on coal substitution and collect estimates of the own-price elasticity of demand for coal and pairwise cross-price elasticities of demand related to coal and three alternative fuels. We provide a comprehensive review and an in-depth analysis of the estimates. We fill in the gap in the literature and address both publication bias and model uncertainty to identify the most important drivers and factors that affect the final elasticities.

Chapter 3

Data

We begin to search for studies reporting the empirical estimates of the elasticity between coal and alternative fuels using Google Scholar. Its algorithm searches for the queried words in full texts of the studies rather than just surveying the title, keywords, or abstract. The query is therefore more precise. As a search query, "interfuel substitution", "fuel substitution" and "energy substitution" is used. We examine the first 500 studies returned by Google Scholar and identify those that may contain empirical estimates of the elasticity. Then, the query is restricted to the most recent studies and the year of the first appearance online is set to 2017 or later. In addition, we inspect all the studies included in the meta-analysis on interfuel substitution conducted by Stern (2012) and download those that include estimates of the elasticity related to coal.

We begin data collection by exploring the studies containing estimates of the elasticity in detail. To be included in our dataset, the study must meet the following inclusion criteria:

- the study must report estimates of the elasticity of demand, as it is the most common unit of measure in the literature concerning interfuel substitution. Since every recalculation of the reported estimate may be to some degree imprecise, this work includes only direct estimates of the elasticity of demand.

- The study must report standard errors or statistics sufficient to compute standard errors. If the study does not report standard errors or other statistics but reports other parameters from which standard errors can be derived, standard errors are computed using the delta method.

Most studies report standard errors directly, but some report t-statistics instead. To compute standard errors, we use the relationship between the t-statistics and the standard error. Few studies report neither the standard error nor the t-statistics. In these cases, it is often possible to calculate standard errors using the delta method. The delta method can be used when the final estimate is a transformation of parameters, for which the standard errors are reported. Then, the standard error of the estimate equals the derivative of the transformation multiplied by the standard error of the corresponding parameter. Taking the derivative of Equation 2.2 and Equation 2.3 with respect to the parameter, this can be formally written as:

$$SE(\eta_{ii}) = \frac{\delta \eta_{ii}}{\delta \gamma_{ii}} SE(\gamma_{ii}) = \frac{\gamma_{ii}}{S_i} SE(\gamma_{ii})$$

$$SE(\eta_{ij}) = \frac{\delta \eta_{ij}}{\delta \gamma_{ij}} SE(\gamma_{ij}) = \frac{\gamma_{ij}}{S_i} SE(\gamma_{ij}),$$

where S_i denotes the cost share of fuel i, and γ_{ii} and γ_{ij} are corresponding parameters from the estimated Equation 2.1.

The final dataset consists of 893 estimates of the own-price elasticity of demand for coal and pairwise cross-price elasticity of demand between coal and three alternative fuels - electricity, gas, and oil. The data are collected from 43 studies. In addition to empirical estimates of the elasticity and their standard errors, we collect 37 variables to account for differences across studies that could possibly capture the heterogeneity in the estimates. These variables are discussed further in detail in Chapter 5.

The oldest study included in the dataset was published in 1976 and the the most recent in 2019. The dataset thus covers four decades of research. Table 3.1 shows the full list of studies included in the meta-analysis.

The full sample is further divided into seven subsets: coal-coal, coal-electricity, coal-gas, coal-oil, electricity-coal, gas-coal, and oil-coal subsets. Because the true effect is expected to vary across different pairs of fuels, we will investigate each own and cross-price elasticity separately. Thus, each subset only contains estimates of the elasticity for a single pair, so that, for example, the coal-coal subset only contains empirical estimates of the own-price elasticity of demand for coal η_{cc} or coal-electricity subset consists of estimates of the cross-price

Table 3.1: Studies used in the meta-analysis

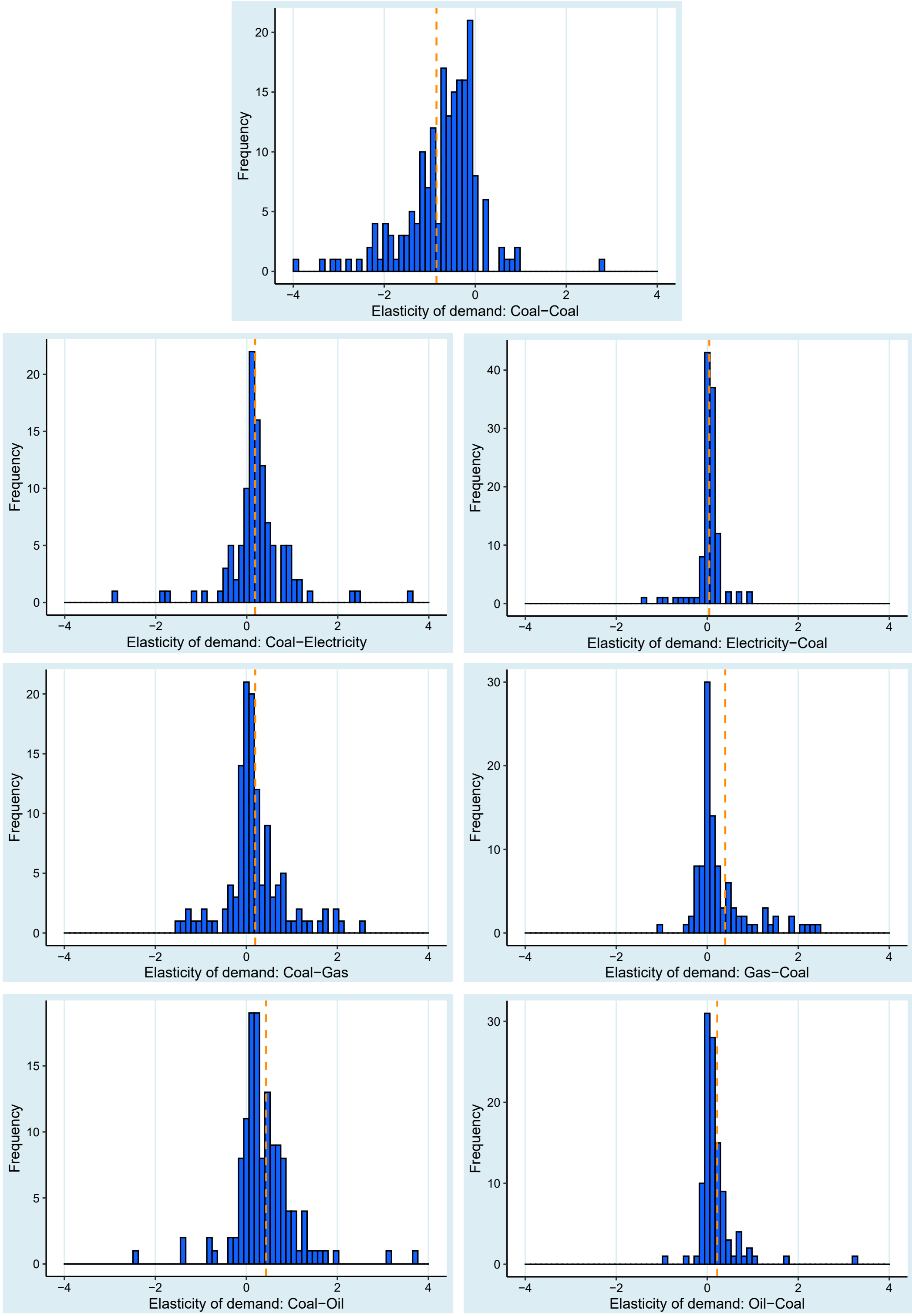
Andrikopoulos et al. (1989)	Halvorsen (1977)
Ma et al. (2009)	Atkinson and Halvorsen (1976)
Harvey and Marshall (1991)	Ma and Stern (2016)
Bello et al. (2020)	He and Lin (2019)
Magnus and Woodland (1987)	Borges and Pereira (1992)
Iqbal (1986)	Pettersson et al. (2012)
Cho et al. (2004)	Jones (1995)
Pindyck (1979)	Considine (2018)
Kim and Labys (1988)	Shahiduzzaman and Alam (2014)
Duncan and Binswanger (1976)	Kim (2019)
Shin (1981)	Fuss (1977)
Ko and Dahl (2001)	Söderholm (2001)
Gri□n (1977)	Li and Lin (2016)
Steinbuks (2012)	Hall (1983)
Lin and Tian (2017)	Steinbuks and Narayanan (2015)
Hall (1986b)	Ma et al. (2008)
Suh (2016)	Hall (1986a)
Wang and Lin (2020)	Taheri (1994)
Urga (1999)	Urga and Walters (2003)
Wang et al. (2019)	Uri (1978)
Uri (1979a)	Wang and Lin (2017)
Yang et al. (2014)	

elasticity of demand between coal and electricity η_{ce}. Figure 3.1 depicts the distribution of the reported elasticities of demand, omitting estimates smaller than -4 and greater than 4 for the sake of readability.[1]

The distributions have their peaks close to zero. More than 92% of the estimates of the own-price elasticity of demand for coal are negative, so the distribution is negatively skewed. This is in line with economic theory, which tells us that the own-price elasticity of demand should take values in the interval ($-\infty$, 0). Such prevalence of the negative estimates could, however, indicate that publication bias is present. Regarding the cross-price elasticity, the distributions are positively skewed, peaking close to zero, implying that the fuels should form weak substitutes. It seems that the substitution possibilities are quite limited, as more than 85% of the estimates lie in the interval (-1,1). Only 8% of the estimates of the coal-electricity elasticity are greater than 1. However, there are no estimates that exceed 1 in the case of the electricity-coal

[1] Table A.1 shows the list of all studies from which the empirical estimates of the elasticity were collected, including those estimates for which it was impossible to calculate standard errors. The total number of collected estimates was 1967 from 65 studies, out of which 893 from 43 studies could be included in the final dataset. Distribution and the summary statistics of all these estimates can be found in the Figure A.1, Table A.2, and Table A.3 in Appendix A

Figure 3.1: Distribution of the reported estimates



Notes: The table shows the distribution of the reported estimates. The vertical line denotes the mean elasticity of demand for each subset. Estimates smaller than -4 and greater than 4 are excluded from the figure for the ease of exposition but included in all statistical tests.

Table 3.2: The mean and median reported elasticity

Subset	Observations	Unweighted mean	Weighted mean	Median
Coal-Coal	193	-0.85	-0.61	-0.56
Coal-Electricity	112	0.19	0.18	0.18
Coal-Gas	124	0.19	0.14	0.10
Coal-Oil	135	0.43	0.30	0.29
Electricity-Coal	114	0.04	0.07	0.05
Gas-Coal	104	0.39	0.33	0.08
Oil-Coal	111	0.22	0.31	0.07

Notes: The table reports the mean and median reported estimate of the elasticity of substitution for each subset. Weighted = estimates are weighted by the inverse of the number of observations reported per study.

Table 3.3: The mean and median standard errors

Subset	Observations	Unweighted mean	Weighted mean	Median
Coal-Coal	193	0.57	0.42	0.19
Coal-Electricity	112	0.24	0.27	0.09
Coal-Gas	124	0.36	0.32	0.09
Coal-Oil	135	0.68	0.61	0.26
Electricity-Coal	114	0.06	0.07	0.03
Gas-Coal	104	0.69	0.57	0.07
Oil-Coal	111	0.79	0.79	0.11

Notes: The table reports the mean and median standard error for each subset. Weighted = estimates are weighted by the inverse of the number of observations reported per study.

elasticity. Around 10% of the coal-gas, gas-coal, coal-oil elasticities and 4% of the oil-coal elasticities are greater than 1.

Table 3.2 shows the unweighted and weighted mean elasticity as well as the median estimate for each subset. Regarding the own-price elasticity of demand for coal, the unweighted mean reported estimate is -0.85, whereas the weighted mean equals -0.61. For all the other elasticities, the mean reported estimate is positive but below 0.5, both weighted and unweighted, suggesting perhaps limited possibilities to substitute between the pairs of fuels. The mean reported estimate attains the lowest value in coal-electricity, coal-gas, and electricity-coal subsets. The median reported estimates are somewhat closer to zero than the means, which are possibly driven by few extreme values. Therefore, the median reported elasticities could be more informative than the mean estimates. The median own-price elasticity for coal is equal to -0.56. The median reported estimate is greatest in the coal-oil and coal-electricity subsets, attaining values of 0.29 and 0.18, respectively. In the remaining subsets, the median elasticity is

Table 3.4: Summary statistics of the reported elasticity

Subset	Observations	St.Dev.	Min	Max
Coal-Coal	193	1.23	-9.69	2.78
Coal-Electricity	112	0.89	-5.29	3.55
Coal-Gas	124	0.66	-1.55	2.52
Coal-Oil	135	0.81	-2.40	5.52
Electricity-Coal	114	0.28	-1.43	0.98
Gas-Coal	104	0.95	-1.06	5.99
Oil-Coal	111	0.58	-0.92	4.36

Notes: The table reports summary statistics of the estimates for each subset. St.Dev. = standard deviation.

Table 3.5: Summary statistics of the standard errors

Subset	Observations	St.Dev.	Min	Max
Coal-Coal	193	1.34	0.0004	13.69
Coal-Electricity	112	0.49	0.001	3.853
Coal-Gas	124	0.90	0.001	7.80
Coal-Oil	135	1.34	0.012	11.72
Electricity-Coal	114	0.07	0.0002	0.32
Gas-Coal	104	3.30	0.0003	24.40
Oil-Coal	111	3.22	0.001	27.08

Notes: The table reports summary statistics of the standard errors for each subset. St.Dev. = standard deviation.

close to zero. The median reported estimates are lowest in the electricity-coal, gas-coal, and oil-coal subsets. Thus, based on the median estimates, it seems that the substitution of other fuels for coal could be slightly more feasible than the reversed substitution of coal for alternative fuels.

Table 3.3 reports the mean and median standard error. Similarly, means are greater than median standard errors, apparently driven by few outliers in the data.

Table 3.4 and Table 3.5 report summary statistics for each subset, including standard deviation and minimum and maximum values of the collected estimates and their standard errors. The own-price elasticity for coal ranges from -9.69 to 2.78. The cross-price elasticity of demand attains values as low as -5.29 in case of coal-electricity elasticity, and as high as 5.99 in case of gas-coal elasticity. Ranging from 0.0002 to 27.08, standard errors seem to vary greatly, too.

There are clearly a few outliers in the data that might dominate the summary statistics. To give them less weight, we winsorize the estimates of the own-price elasticity for coal, coal-oil, gas-coal, and oil-coal elasticity. Using this technique, the estimates do not have to be excluded from the dataset, but instead, they are replaced by more plausible values based on the level of winsorization. This can be formally written as follows:

$$\eta_w = min[max(\eta, \eta_p), \eta_{(1-p)}], \tag{3.1}$$

where p denotes the percentage of data we want to replace from each side, η_p is the computed percentile representing the lower threshold for estimates of elasticity and $\eta_{(1-p)}$ represents the upper threshold for estimates. Similarly, I winsorize the standard errors in coal-coal, coal-oil, gas-coal, and oil-coal subsets.

$$SE(\eta)_w = min\{max[SE(\eta), SE_p], SE_{(1-p)}\}, \tag{3.2}$$

where SE_p is the computed percentile representing the lower threshold for the reported standard errors and $SE_{(1-p)}$ denotes the upper corresponding threshold (Miller, 1993). We winsorize at 1% level. In this case, p is equal to 0.01, or, in other words, we compute 1^{st} and 99^{th} percentile for the elasticity and the standard errors.

Additionally, Figure A.2 and Figure A.3 in Appendix A show that the estimates vary both across and within studies. The potential sources and in-depth analysis of heterogeneity will be further discussed in Chapter 5, while the following Chapter 4 examines publication bias, which could also account for some of the differences among the estimates.

Chapter 4

Publication bias

4.1 Defining publication bias

Publication bias, also known as selective reporting, arises when certain results become reported with systematically higher probability. Consequently, substantial reported evidence gathered from studies might be distorted and biased representation of actual knowledge. As Rosenthal (1979) points out, we can never be sure how many studies have never been reported and refers to this problem as "the file drawer problem", because part of the evidence remains hidden.

Ideally, all the estimates should be reported and studies should be chosen for publication based on their quality. However, researchers sometimes tend to report only those results which are intuitive and consistent with theory or previous research, followed by publishers who often dislike studies reporting findings different from existing evidence (Thornton and Lee, 2000). Clearly, imprecise and faulty results might emerge due to noise in the data or the choice of methodology. Naturally, it often makes sense to discard such estimates and instead, try to bring the results closer to the true effect. However, noise in the data or improper methodologies can likewise produce exaggerated estimates. These are, unlikely obviously incorrect and counter-intuitive estimates, much harder to spot. If seemingly faulty estimates are discarded but overestimated ones are reported, publication bias arises (Gechert et al., 2021). This is called type I publication bias.

Similarly, studies might be less likely published and results less likely disclosed if researchers find no supportive evidence of the effect and instead, they rather discover nonexistence of it. If the authors prefer to present only statistically significant results, type II publication bias emerges (Stanley, 2005).

Such selective reporting can lead to distorted results and incorrect conclusions about the true effect. In our case, economic theory does not give much support to positive own-price elasticity for coal. By definition, if the own-price elasticity lies in the interval (-∞, 0), the increase in price of coal should result in lower demand for coal, whereas positive own-price elasticity would imply that increase in price of coal leads to higher demand. Similarly, coal and alternative fuels might be intuitively assumed to form substitutes rather than complements, or, in other words, the cross-price elasticity might be assumed to be positive rather than negative. Therefore, researchers could be tempted to under-report results not in line with the theoretical background and intuition. On these assumptions, positive own-price elasticity and negative cross-price elasticity might be under-reported. Hence, we have to take into account that publication bias might be present.

Stern (2012) provides a short discussion on publication bias in the literature concerning interfuel substitution, suggesting that while he does not suspect the significance of the results to be of much importance, positive own-price elasticities might be censored. In other words, he argues that even though type II publication bias might be absent, type I publication bias could possibly emerge. He uses the inverse of the square root of the sample size used to generate estimates as a proxy for publication bias in the meta-regression analysis. As noted by Begg and Berlin (1988), publication bias should be proportional to the inverse of the square root of sample size. Additionally, Stanley (2008) argues that if we include this variable as a proxy for publication bias in the meta-regression, the estimated intercept will represent the value of the elasticity generated using an infinite sample size. This way, the elasticity is corrected for publication bias. This approach is followed by Chen (2017), but both of them find that the coefficient of proxy for publication bias is insignificant in most cases. However, to our knowledge, there is no better analysis of publication bias in the literature on fuel substitution. We now turn out attention to different techniques used to detect selective reporting.

4.2 Testing for publication bias

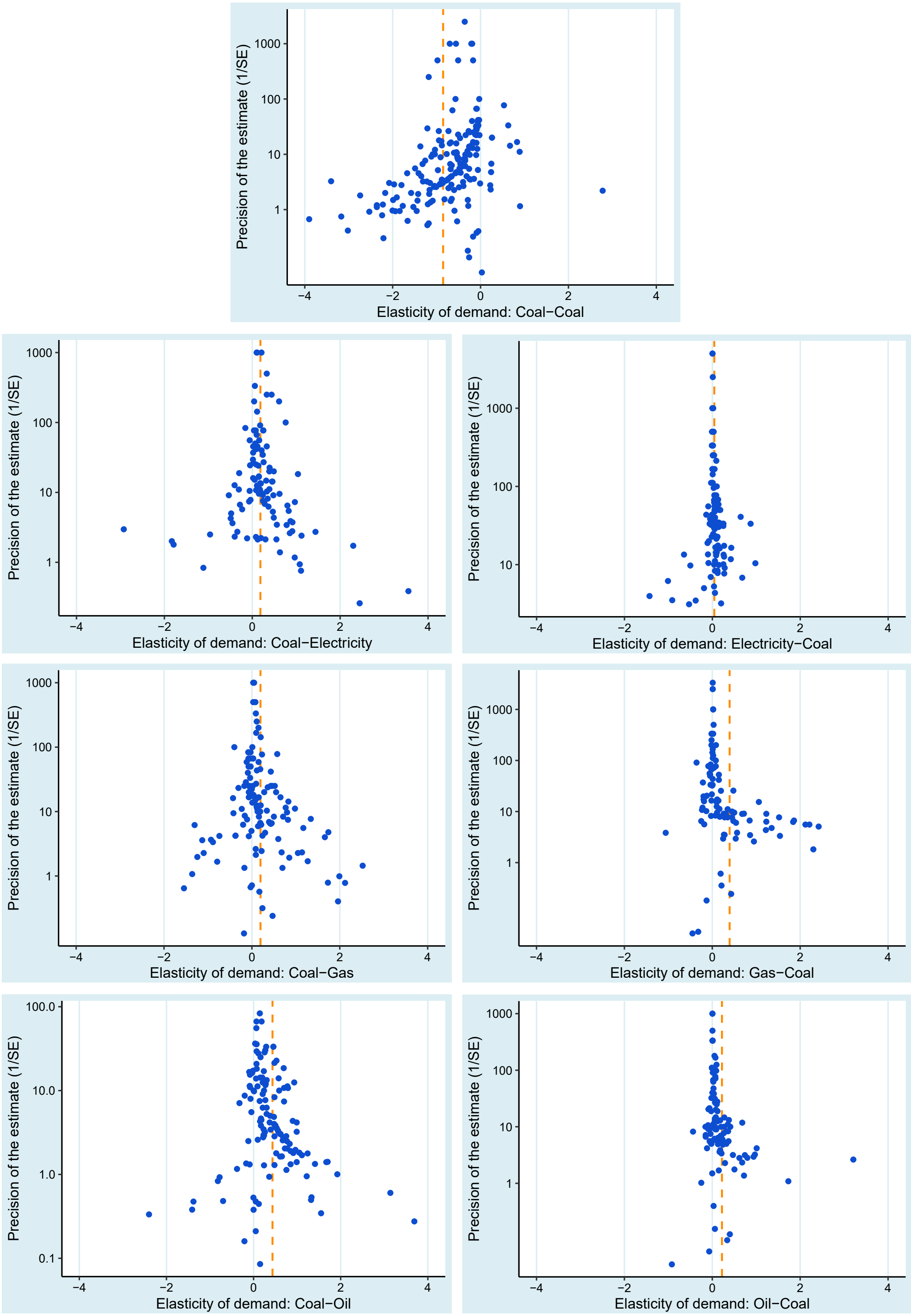
The most simple graphical test of publication bias is the funnel plot. Introduced by Egger et al. (1997), the original funnel plot was a plot of the effect estimates against the sample size used to generate these estimates as a proxy for precision. Another proxy for precision is the inverse of the standard errors of the reported estimates, which is the proxy we use. If publication bias is absent, the estimates should be symmetrically dispersed around the true effect (Duval and Tweedie, 2000). Figure 4.1 depicts funnel plots for each subset. Funnel plots are constructed using original data, not the winsorized ones. The vertical line denotes the mean and estimates smaller than -4 and greater than 4 are excluded from the figure for the sake of readability.

The constructed funnels are not exactly symmetrical. In the coal-coal subset, most estimates seem to lie in the interval (-1,0), while there are only a few reported positive estimates, resulting in an asymmetrical funnel plot. In the gas-coal and oil-coal subsets, there are apparently fewer negative estimates of the elasticity. Similarly, there might be less negative estimates of the electricity-coal elasticity. In the coal-oil subset, it seems that there could be slightly more estimates on the left side of the mean than on the right side and the funnel plot seems somewhat asymmetrical. On the contrary, the estimates of the coal-electricity and coal-gas elasticity seem to be dispersed relatively symmetrically around the mean. Nonetheless, based on such simple plots, we can not be certain. To wrap up, there seems to be some evidence of publication bias, although we suspect it might not affect all the subsets.

However, as noted by Stanley and Doucouliagos (2010), funnel plots are not sufficient tests by themselves. There are other objective regression-based tests that can be used to verify the subjective interpretation of simple funnel plots. In addition, heterogeneity among estimates might also complicate the inspection of such simple graphical tests as it could cause slight asymmetry, too. Thus, we now continue with several linear techniques to detect publication bias.

Regression-based techniques can be considered more objective and rigorous approach compared to funnel plots. The tests are based on the assumption that if publication bias is absent, the estimates and standard errors should be statistically independent of each other. The equation can be written as:

Figure 4.1: The funnel plots suggest some publication bias



Notes: The figure depicts funnel plot for each subset. In the absence of publication bias, estimates should be symmetrically distributed around the mean, which is denoted by the vertical line. Estimates smaller than -4 and greater than 4 are excluded for ease of exposition but included in all statistical tests.

$$\eta_{ij} = \eta_0 + \gamma SE(\eta_{ij}) + \epsilon_{ij}, \tag{4.1}$$

where η_{ij} denotes the i^{th} estimated elasticity of demand from the j^{th} study, SE(η_{ij}) is its standard error, γ stands for the intensity of publication bias, ϵ_{ij} for the error term and as noted by Stanley (2005), η_0 is the mean elasticity corrected for publication bias.

Testing for the significance of γ can thus be viewed as a statistical test for the presence of publication bias, also called FAT or Funnel asymmetry test. Sign of the coefficient and magnitude can be interpreted as the direction and intensity of publication bias. Testing for statistical significance of the estimated intercept can be considered a test of the existence of true effect different from zero, called PET or Precision-effect test. If we, for example, do not reject the null hypothesis $H_0 : \eta_0$=0, we can assume that the true elasticity is in fact 0.

However, estimates are unlikely to be independent within studies. Then, the usual assumption that ϵ_{ij} is identically and independently distributed and follows standard normal distribution is violated. Violation of this assumption does not imply that the estimated coefficients will be biased, but the estimated standard errors might be imprecise. Therefore, the inference based on statistical tests could be invalid. In order to avoid this, we cluster the standard errors at the study-level. This approach allows for correlation between standard errors within individual studies but assumes independence across studies.

Firstly, we estimate the between-effects model. BE uses between-study variance and allows for more balanced study weights. In this specification, large studies are assigned less weight than the small ones. Because some of the reviewed studies report only a few estimates, while other disclose a large number of the estimated elasticities, we believe that this approach is appropriate. Thus, we estimate the BE model to treat for variation in the study size.[1] In addition to the between-effects method, we also use weighted-least squares. Because extreme estimates are often linked to high standard errors, heteroskedasticity might arise. Stanley and Doucouliagos (2017) argue that using the inverse of

[1] It is also common to estimate the fixed-effects model that incorporates study-specific terms to account for features that vary across studies. However, we find the BE specification more suitable. The FE model would be based on the studies with a large number of estimates, which could be problematic and counter-intuitive.

standard error as weights should treat the heteroskedasticity in the regression. This way, imprecise estimates are given less weight than the more precise ones. The equation to be estimated is

$$\frac{\eta_{ij}}{SE(\eta_{ij})} = \gamma + \eta_0 \frac{1}{SE(\eta_{ij})} + \quad_{ij}$$

and interpretation of the coefficients becomes reversed. The estimated intercept γ measures the intensity of publication bias and the coefficient η_0 represents the mean elasticity corrected for bias. As an alternative, we weight Equation 4.1 by the inverse of the number of estimates reported per study. In this specification, every study is assigned the same weight. Lastly, we will address potential endogeneity of the standard errors and run an instrumental variable regression. So far, it was assumed that the standard errors are exogenously given, but such assumption might be faulty. Standard errors and estimates can be jointly determined by the design of a particular study. An intuitive instrument would be the inverse of the square root of the sample size used to generate estimates. While the root is correlated with standard errors, it is unlikely to be correlated with the choice of methodology and estimation technique and thus, should correct for possible endogeneity.

Table 4.1 presents the results from all these regressions. In the case of coal-coal, coal-oil, gas-coal, and oil-coal subsets, we use data winsorized at 1% level at both distribution's sides. To interpret the results, we will follow the classification provided in Doucouliagos and Stanley (2013). If $|\gamma| < 1$ or if the coefficient is insignificant, publication bias is small or modest. If $1 < |\gamma| < 2$, then the evidence suggests that publication bias is substantial. If $|\gamma| > 2$, the publication bias is severe.

Table 4.1: Linear techniques suggest some publication bias

Coal-Coal subset	BE	Precision	Study size	IV
Standard error	-1.2450***	-0.4072	-0.5879***	-1.3888***
(Publication bias)	(0.1709)	(1.5746)	(0.0693)	(0.3553)
Constant	-0.3533***	-0.4282***	-0.1701	-0.1139
(Mean beyond bias)	(0.0538)	(0.0823)	(0.1084)	(0.2189)
Observations	193	193	193	193
Coal-Electricity subset				
Standard Error	-0.1954	1.5826	0.4443*	0.3947
(Publication bias)	(0.2603)	(1.2938)	(0.2178)	(0.6845)
Constant	0.1978***	0.1609***	0.0937	0.0899
(Mean beyond bias)	(0.0496)	(0.0302)	(0.0582)	(0.1101)
Observations	112	112	112	112
Coal-Gas subset				
Publication bias	0.2948	1.2977	0.1744**	0.3512
(Standard error)	(0.2353)	(0.6818)	(0.0628)	(0.8443)
Effect beyond bias	0.1191*	0.0459***	0.0369	0.0696
(Constant)	(0.0510)	(0.0077)	(0.0684)	(0.2667)
Observations	124	124	124	124
Coal-Oil subset				
Publication bias	0.5920***	0.9950***	0.3487***	-0.6109
(Standard error)	(0.1530)	(0.2023)	(0.0552)	(1.1616)
Effect beyond bias	0.2114***	0.1411***	-0.0895	0.8094
(Constant)	(0.0363)	(0.0268)	(0.0770)	(0.7163)
Observations	135	135	135	135
Electricity-Coal subset				
Publication bias	-0.9396*	2.6314***	0.1930**	1.9978
(Standard error)	(0.3907)	(0.5302)	(0.0584)	(2.2565)
Effect beyond bias	0.1038***	0.0036***	-1.4994	-0.0666
(Constant)	(0.0260)	(0.0010)	(0.9872)	(0.1327)
Observations	114	114	114	114
Gas-Coal subset				
Publication bias	1.7614***	1.7317**	0.4191***	0.3733
(Standard error)	(0.1084)	(0.6350)	(0.0878)	(0.5884)
Effect beyond bias	0.0076***	0.0101***	-0.0288***	0.1435
(Constant)	(0.0008)	(0.0008)	(0.0070)	(0.2779)
Observations	104	104	104	104

Continued on next page

Table 4.1: Linear Techniques suggest some publication bias (continued)

Oil-Coal subset	BE	Precision	Study size	IV
Publication bias	1.0055***	1.5099***	0.3212***	0.1744
(Standard error)	(0.1593)	(0.2225)	(0.0939)	(0.2259)
Effect beyond bias	0.0222	0.0044**	-0.0094	0.0942
(Constant)	(0.0146)	(0.0014)	(0.0166)	(0.1036)
Observations	111	111	111	111

Notes: The table reports results of the regression $\eta_{ij} = \eta_0 + \gamma SE(\eta_{ij}) + {}_{ij}$. η_{ij} and SE(η_{ij}) stand for the i^{th} estimate of the elasticity of demand and its standard error from the j^{th} study. BE = between-effects model. Precision = the inverse of the standard error of the reported estimate is used as the weight. Study size = the inverse of the number of estimates reported per study is used as the weight. IV = the inverse of the square root of the sample size used to generate the estimate is used as an instrument for the standard error. The standard errors are clustered at the study level and reported in parentheses. ***, ** and * denote statistical significance at 1%, 5% and 10% level.

In the case of the own-price elasticity for coal, the estimated coefficient measuring the intensity of bias is larger than 1 based on the BE model. According to Doucouliagos and Stanley (2013), this would imply that the bias is substantial. The statistically significant estimates of the mean beyond bias are approximately -0.40 – two times smaller than the mean reported estimate. There seem to be only little publication bias present in the estimates of the coal-electricity elasticity. Besides, the estimated mean beyond bias is close to the mean reported estimate. BE and weighted-by-precision specifications estimated it to be on average 0.18. Similarly, the estimates of the coal-gas elasticity do not seem to suffer from substantial publication bias. The estimated coefficients measuring the intensity of bias are small or insignificant. The estimated mean beyond bias is 0.12 based on the BE model. According to the weighted-by-precision specification, it equals approximately 0.05. Neither is publication bias estimated to be substantial in the coal-oil subset based on the BE specification. The statistically significant estimated mean ranges between 0.14 and 0.21. In the electricity-coal subset, the publication bias is severe according to the weighted-by-precision specification. Based on the BE model, the estimated coefficient measuring the intensity of bias is statistically significant and very close to one in the absolute value. The mean beyond bias is 0.10 according to the BE model, although the weighted-by-precision estimated it to be almost 0. Publication bias seems to substantially affect also the estimates of the gas-coal elasticity. The coefficients representing bias are statistically significant and larger than 1 according to both the BE model and weighted-by precision model. The mean beyond bias is estimated to be very close to zero. Lastly, the bias is substantial in the oil-coal subset, too, based on the BE and weighted-by-precision models. The mean beyond bias is very close to zero as well.

However, we note that the weighted-by-studysize specification did not detect substantial publication bias in any subset. The estimated coefficient capturing bias is the largest and statistically significant in the coal-coal subset, but still, it is smaller than 1. In all the other subsets, it is either smaller or insignificant. Thus, after giving each study the same weight, it seems that in cases where publication bias is present, it is mainly driven by a few studies.

Regarding the instrumental variable regression, the regression results are not statistically significant. One exception is the coefficient representing publication bias in the coal-coal sample. However, it is basically impossible to draw any further conclusions based on this specification. The inverse of the square root of the sample size is found to be a weak instrument and the estimated coefficients are highly uncertain. Still, we decided to report the results of the regression to demonstrate the attempt to account for endogeneity. Several different transformations of the sample size were also used as an instrument, namely the inverse of the sample size, the inverse of the sample size squared, and the logarithm of the sample size. Nonetheless, none of these instruments seemed to be efficient as statistical tests declared that all these transformations of the sample size are weak instruments. Estimated coefficients and intercepts remained imprecise.

All the aforementioned tests assumed that publication bias is a linear function of the standard error. However, this assumption might be violated if we expect to find non-linear breaks at some data points. To take this into account, we adopt several non-linear methods.

The first non-linear method we use is the so-called Weighted average of adequately powered. This method accounts for the tendency to publish only significant estimates, those that pass the 1.96 t-statistics threshold. The focus of the test is thus on the estimates with adequate statistical power, that is, above 80%, and they are chosen based on their standard errors (Ioannidis et al., 2017).

The second non-linear specification that we use is the Stem-method, developed by Furukawa (2019). This method adopts the logic of the Top10 test (Stanley et al., 2010). Top10 method is based on the assumption that if estimates with higher statistical significance are more likely to be published than the insignificant ones, the substantial evidence gathered from studies will be biased. Stanley et al. (2010) argue, that the meta-analyst might as well just exclude 90% of the data and only keep the estimates with the highest precision. The stem-based method, on the other hand, exploits the trade-off between variance and bias and suggests using a certain proportion of the most precise estimates, actually, the stem of the funnel plot. The number of estimates included is determined by minimizing the mean squared error based on the bias and variance. As we increase the number of estimates included, the bias increases too. On the other hand, this is partially offset by decreasing variance due to more information. Formally speaking,

$$minMSE(n) = Bias^2(n) + Var(n), \tag{4.2}$$

where MSE is the mean squared error and Var denotes the variance.

Lastly, we follow Andrews and Kasy (2019) and use the Selection model. They demonstrate how the known conditional publication probability can be used to correct for publication bias. The Selection model applies a different weighting scheme to the imprecise estimates, rather than simply discarding them. A step function assigns different weights to every interval of the reported p-values. We set the cut-offs by looking at the histograms of t-statistics, which can be found in Figure B.1 in Appendix B. However, a crucial but strong assumption of this model is that the estimates and their standard errors should be statistically independent in the absence of publication bias, or, in other words, the correlation between the estimates and standard errors should be equal to zero. It is not unlikely that such assumption does not hold, as the estimates could be correlated with standard errors because of the study design or data used, and we decide to test this assumption. Since we do not know the distribution of the estimates and the standard errors were the publication bias absent, it is suggested that under the null hypothesis that all assumptions of the Selection model hold, we should be able to estimate the correlation between the two by weighing the estimates by the inverse of the estimated publication probabilities (Kranz, 2022). The results of this estimation are presented in Table B.1 in Appendix B, indicating that the correlations between estimates and their standard errors are not, in fact, zero. Therefore, the key assumption seems to be violated and the Selection model, although it has the most rigorous foundation, should be interpreted only with caution.

Furthermore, we also intended to use the caliper test, developed by Gerber and Malhotra (2008). In addition to the IV regression, this test would also allow us to relax the exogeneity assumption. While the caliper test does not assume any relationship between the estimate and its standard error, it can be used to compare and test the statistical significance of the relative frequency of the t-statistics below and above a certain threshold. If publication bias is absent, the number of t-statistics below the set threshold should be approximately equal to the number of t-statistics above this threshold. On the other hand, if the latter outnumber the former and this difference is statistically significant, publication bias is most likely to be present. Unfortunately, due to a low number of observations dispersed around the suspicious breaks, we had to either set the caliper width to be too wide or, on the other hand, we had too few observations. It is thus almost impossible to draw any conclusions based on this test and it would have to be interpreted with extreme caution. The results of the tests can be found in Table B.2 and Table B.3 in Appendix B. There are a few additional methods that do not rely on the exogeneity assumption between elasticity estimates and their standard errors. Another technique, called p-uniform*, developed by van Aert and van Assen (2021), is based on the statistical principle that the p-values should be uniformly distributed at the true effect. To estimate the mean corrected elasticity, it searches for a number that approximates the uniform distribution of p-values. However, our data were not fit for this technique as the test failed to find such a number, and thus, we do not disclose the results of this method. Moreover, Elliott et al. (2022) introduced two novel histogram-based tests. Unfortunately, because the estimation is driven by the number of bins the researcher uses to test for the publication bias, we were unable to use this technique as we did not have enough observations. Thus, these specifications will not be further discussed, but nonetheless, we mention the attempt to control for potential endogeneity of the standard errors.

Table 4.2: Non-linear techniques suggest some publication bias

Coal-Coal subset	WAAP	Stem-method	Selection model
Mean beyond bias	-0.3780***	-0.4167*	-0.2944
	(0.0465)	(0.1518)	(0.2300)
Observations	193	193	193
Coal-Electricity subset			
Effect beyond bias	0.1830***	0.2262***	0.0700
	(0.0329)	(0.0364)	(0.0515)
Observations	112	112	112
Coal-Gas subset			
Mean beyond bias	0.0540*	0.0445**	0.0687
	(0.0322)	(0.0202)	(0.1182)
Observations	124	124	124
Coal-Oil subset			
Mean beyond bias	0.2120***	0.1697***	0.1734***
	(0.0384)	(0.0576)	(0.0000)
Observations	135	135	135
Electricity-Coal subset			
Mean beyond bias	0.0141***	0.0053	0.0019
	(0.0046)	(0.0063)	(0.0033)
Observations	114	114	114
Gas-Coal subset			
Mean beyond bias	0.0115*	-0.0011	-0.0131**
	(0.0063)	(0.0050)	(0.0055)
Observations	104	104	104
Oil-Coal subset			
Mean beyond bias	0.0275***	0.0043	0.0044
	(0.0087)	(0.0118)	(0.0129)
Observations	111	111	111

Notes: The table shows results of 3 non-linear techniques to detect publication bias. WAAP model developed by Ioannidis et al. (2017) is based on the estimates with adequate power. Stem-method introduced by Furukawa (2019) focuses on certain proportion of the estimates, determined by minimizing the mean squared error based on the bias and variance. The Selection model by Andrews and Kasy (2019) estimates the corrected mean by assigning a different weighting scheme to each interval of the reported p-values. The standard errors are reported in parentheses. For the WAAP method, we cluster standard errors at the study level. ***, ** and * denote statistical significance at 1%, 5% and 10% level.

Table 4.2 reports the results from non-linear tests. In the coal-coal, coal-oil, gas-coal, and oil-coal subsets, the tests were performed using data winsorized at 1% level. The estimates are mostly consistent with the estimates from linear regressions. The mean corrected own-price elasticity of demand for coal is equal to -0.38, based on the WAAP method. In the coal-electricity subset,the WAAP method estimated the mean corrected for bias to be 0.18. Regarding the elasticity between coal and gas, the effect beyond bias is approximately 0.05, in line with the previously used weighted-by-precision specification. The estimated mean corrected coal-oil elasticity ranges between approximately 0.17 and 0.21. In the three remaining cases, the mean corrected electricity-coal, gas-coal, and oil-coal elasticities are estimated to be almost zero.

To conclude, there seems to be some evidence that publication bias is present. Publication bias seems to substantially affect the estimates of the own-price elasticity for coal, electricity-coal, gas-coal and oil-coal elasticity. However, after giving each study the same weight, the bias became very modest. Thus, it seems that the bias might be driven by a few specific studies.

The estimated means beyond bias from both the linear and non-linear tests are mostly in line. The mean corrected own-price elasticity for coal is approximately -0.40. The estimated mean coal-electricity elasticity is on average around 0.19. Mean corrected coal-gas elasticity ranges between 0.05 and 0.12 and the mean corrected coal-oil elasticity ranges between 0.14 and 0.21. Although the estimated means are relatively small, we can not reject the alternative hypothesis: $H_0 : \eta_0 = 0$. On the other hand, the mean corrected electricity-coal, gas-coal, and oil-coal elasticities are consistently estimated to be almost zero. In these cases, we can not reject the null hypothesis $H_0 : \eta_0 = 0$. Our findings indicate that the substitution possibilities are, in general, very limited. These results also support our previous suggestions that the substitution of other fuels for coal could be somewhat more feasible than the reversed substitution of coal for alternative fuels.

Nonetheless, we have to bear in mind that there are other drivers possibly affecting the relationship between the estimates and their standard errors. Heterogeneity across studies could also impact the results of the tests used to detect publication bias. In Chapter 5, we provide detailed discussion of possible explanations behind this heterogeneity.

Chapter 5

Why do estimates differ?

Besides testing for publication bias, this work will also account for other potential drivers of the true effects and sources of heterogeneity. For example, in previously conducted meta-analysis, Stern (2012) finds that the estimates of the elasticity are greater at the industry-level than at the state-level. Similarly, estimates produced by cross-sectional or panel data tend to be greater than the elasticities generated by time-series (Bacon and Mundial, 1992; Stern, 2012). Additionally, Chen (2017) finds that substitutability between fuels in the electricity generation sector is greater for the US, compared to other selected countries and regions, including Europe, Mexico, Turkey, Japan, and Australia.

Nonetheless, many other characteristics could influence the substitutability between coal, electricity, gas, and oil as well as the reported estimates of the elasticity and we will now continue with a detailed investigation of these variables. The regression to analyze the impact of all these variables on the elasticity of demand will take the following form:

$$\eta_{ij} = \eta_0 + \beta_j \sum_j X_{ij} + u_{ij}, \tag{5.1}$$

where η_{ij} denotes the i^{th} estimate from the j^{th} study, η_0 stands for the constant term, β_j for the vector of coefficients, X_{ij} for the vector of explanatory variables and u_{ij} is the error term.

Altogether, we collected 41 variables that could explain the heterogeneity among estimates. However, many of these variables are correlated with each other, possibly coding the same information. Hence, after inspecting multicollinearity, we decided to only use 26 out of them in the final regression.

Moreover, in some subsets, we had to additionally merge or drop some of the variables due to persisting collinearity or lack of observations in certain categories, resulting in fewer explanatory variables. For the sake of simplicity, we divide all the potential drivers into several categories. Namely, we discuss few specifications such as endogeneity controls, data characteristics, possible structural variation, design of the production function, estimation techniques, and lastly, publication characteristics. Table 5.1 reports the simple means for each subset.

Table 5.1: Description and summary statistics of regression variables

Variable	Description	C-C Mean	C-E Mean	C-G Mean	C-O Mean	E-C Mean	G-C Mean	O-C Mean
Elasticity estimate	Estimate of the elasticity of demand (dependent variable).	-0.83	0.19	0.19	0.43	0.04	0.39	0.21
Standard error (SE)	Standard error of the estimated elasticity.	0.52	0.24	0.36	0.63	0.06	0.66	0.67
Specifications								
Control variables	=1 if study used proxies, dummy control variables or seasonal dummy variables in the regression to correct for potential endogeneity.	0.21	0.25	0.29	0.16	0.25	0.35	0.20
Technological change	=1 if study captured technological change in the model.	0.51	0.53	0.62	0.43	0.53	0.56	0.33
Data characteristics								
Time-series data	=1 if study used time-series data to generate the estimates.	0.52	0.56	0.43	0.60	0.57	0.50	0.70
Panel data	=1 if study used panel data to generate the estimates	0.43	0.38	0.51	0.34	0.37	0.42	0.23
Cross-sectional data	=1 if study used cross-sectional data to generate the estimates.	0.05	0.06	0.06	0.06	0.06	0.08	0.07
High frequency	=1 if study used quarterly or monthly data to generate the estimates.	0.05	0.04	0.03	0.05	0.04	0.02	0.04
Annual frequency	=1 if study used yearly data to generate the estimates.	0.82	0.94	0.80	0.77	0.94	0.92	0.91

Continued on next page

Table 5.1: Description and summary statistics of regression variables (continued)

Variable	Description	C-C Mean	C-E Mean	C-G Mean	C-O Mean	E-C Mean	G-C Mean	O-C Mean
Low frequency	=1 study used 5-year data to generate the estimates.	0.11	0.00	0.14	0.16	0	0	0
Elasticity year	The year for which the elasticity was estimated minus the earliest year for which the elasticity was estimated.	33.78	36.75	36.01	32.33	36.86	37.58	33.35
Data length	The number of years captured by the data used to generate the estimates.	18.35	20.22	19.24	18.47	20.24	20.03	19.08
Structural variation								
GDP per capita	Logarithm of the GDP per capita of the country for which the elasticity was estimated, measured at the midyear of the data used to generate the estimates.	9.29	9.62	9.92	9.27	9.60	9.99	9.32
Macro-level	=1 if the elasticity was estimated for the whole country.	0.34	0.32	0.31	0.39	0.32	0.18	0.29
Industrial sector	=1 if the elasticity was estimated for the industrial or manufacturing sector.	0.29	0.46	0.40	0.18	0.46	0.47	0.21
Subindustrial sector	=1 if the elasticity was estimated for the subindustrial or submanufacturing sector	0.26	0.22	0.14	0.32	0.22	0.16	0.37
Electric power generation sector	=1 if the elasticity was estimated for the electric power generation sector.	0.11	0.00	0.15	0.11	0.00	0.18	0.14

Continued on next page

Table 5.1: Description and summary statistics of regression variables (continued)

Variable	Description	C-C Mean	C-E Mean	C-G Mean	C-O Mean	E-C Mean	G-C Mean	O-C Mean
Design of the production function								
Translog framework	=1 if the model specification used for estimation is the translog function.	0.84	0.75	0.75	0.90	0.75	0.70	0.87
Logit model	=1 if the model specification used for estimation is the linear logit model.	0.14	0.24	0.22	0.08	0.24	0.26	0.10
Different model	=1 if the model specification used for estimation is different from translog or linear logit model.	0.02	0.01	0.03	0.02	0.01	0.04	0.03
Non-homothetic model	=1 if the model used for estimation is non-homothetic.	0.03	0.06	0.07	0.00	0.07	0.07	0.00
Estimation technique								
SUR method	=1 if the seemingly unrelated regression or its variations (iterative Zeller's efficient method) is used.	0.74	0.71	0.70	0.72	0.71	0.64	0.66
ML method	=1 if maximum-likelihood or its variations (non-linear ML or full-information ML) is used.	0.20	0.19	0.23	0.22	0.18	0.28	0.27
Different method	=1 if estimation method different from SUR or ML is used.	0.06	0.11	0.06	0.06	0.11	0.08	0.07
Dynamic model	=1 if the model specification used for estimation is dynamic	0.21	0.35	0.28	0.15	0.34	0.34	0.18

Continued on next page

Table 5.1: Description and summary statistics of regression variables (continued)

Variable	Description	C-C Mean	C-E Mean	C-G Mean	C-O Mean	E-C Mean	G-C Mean	O-C Mean
Long-run e□ect	=1 if a long-run e□ect was estimated.	0.09	0.15	0.15	0.05	0.15	0.17	0.06
Publication characteristics								
Impact factor	The recursive discounted impact factor from RePEc	0.35	0.30	0.44	0.43	0.30	0.32	0.31
Citations	The logarithm of the number of citations of the study in Google Scholar per year since the study was published.	0.91	0.92	0.95	0.86	0.91	0.90	0.84

Notes: C=coal, E=electricity, G=gas, O=oil. The pair of letters denotes the subset containing corresponding cross-price elasticities between the two fuels. SE = standard error, GDP per capita = gross domestic product per capita, measured at constant 2015$, SUR = seemingly unrelated regression, ML = Maximum-likelihood.

5.1 Variables

Specifications

This category includes study settings and control variables that are used by researchers to correct for endogeneity. Because controlling factors to account for seasonality, shocks or other potential impacts used by authors are particularly study-specific, we merge these studies into one category and create a dummy equal to one if a study uses any control variables. To be specific, Steinbuks (2012) and Steinbuks and Narayanan (2015) use structural shift variables to capture unexplained shifts in fuel consumption. Besides, Steinbuks and Narayanan (2015) and Urga and Walters (2003) include the output in their models to account for unobserved changes in the economy which could possibly affect fuel consumption. Steinbuks (2012), who uses panel data for 15 manufacturing industries to estimate the elasticity of demand, introduces a dummy variable for those industries that use combined heating and power systems, because these are expected to be more efficient in fuel use and thus, should produce greater elasticities of demand. Pettersson et al. (2012) investigate fuel substitution in the European power generation sector and use a dummy variable equal to one for countries in which baseload purposes are heavily served by non-fossil sources and zero otherwise. This allows for testing whether the responses of the two groups of countries to changes in fuel prices are equal or not. In addition, they use a variable denoting the net generation of nuclear energy and hydropower for each country. This way, the problem of potential bias arising from omitting the role of nuclear and hydropower in relevant countries is addressed. Furthermore, they use a dummy variable equal to one for the post-liberalization years in each country included in their dataset to capture the effects of electricity deregulation on fuel choice. Uri (1978) introduces a degree day variable to take into account natural gas curtailments. Heating degree day is a unit that accounts for days when the outdoor mean temperature falls below a certain specified temperature, in this case, 65°F or, approximately, 18°C. As argued by Uri (1978), this variable is chosen because gas curtailments are more likely to happen during the heating season and especially during more severe winters. Finally, Steinbuks and Narayanan (2015), who distinguish between fossil fuel producing and other countries to investigate fuel substitution, use a natural resource endowment as an instrumental variable to correct for potential endogeneity of fuel production. They argue that the natural endowment measures the feasibility of fossil fuel production, which is highly correlated with

the actual fossil fuel production. Yet, it should be uncorrelated with other variables affecting fuel production such as government subsidies or regulations. Therefore, fossil fuel production instrumented by natural resource endowment should not suffer from endogeneity.
Besides different control variables, some researchers also included seasonal controls. For example, Kim (2019) applies monthly dummies or Cho et al. (2004) quarterly dummies to account for seasonal variations. Additionally, Pettersson et al. (2012) and Söderholm (2001) include a dummy variable for miner's strike of 1984–1985 within the British coal industry.

Another important variable is technological change and therefore, we create a dummy variable equal to one if the study modeled technological change. As Stern (2012) argues, it is reasonable to assume that technological change is non-neutral. Because technological change might influence the ease with which it is possible to substitute between different fuels and affect the efficiency of energy use, ignoring technological progress could introduce bias. Thus, under the assumption that technological change is correlated with prices for energy inputs, probably the most important omitted variable in the literature is the state of technology. However, only around 50% of the studies modeled technological change. Following Welsch and Ochsen (2005), most researchers simply use a linear time trend to capture the state of technology. This approach is followed by, for example, Ma et al. (2008), Shahiduzzaman and Alam (2014), or Lin and Tian (2017). A slightly different approach is chosen by Harvey and Marshall (1991), who correct for technological biases by including stochastic instead of deterministic trend components. Nonetheless, we note that modeling technological change is not a trivial exercise. Even though the literature concerning interfuel substitution does not seem to be disturbed by how technological change should be captured, this issue is illustrated in, for example, Borjas et al. (2012). This paper from the field of labour economics demonstrates how including time trend in different transformations affects the final estimates. Thus, one has to make an assumption about which transformation is the best fit. As they note: "different assumptions yield different conclusions," (Borjas et al., 2012, p. 209). We have to bear in mind that it is uncertain how the time trend should be included in the model. Incorrect modeling could tweak the final estimates and it might be hard to assess the impact of this variable.

Data characteristics

This category comprises information regarding the data used by reviewed studies. Firstly, we control for data dimension. Bacon and Mundial (1992) finds that using cross-sectional data leads to greater estimates of elasticities, while time-series data often produce smaller estimates. Further evidence is provided by Stern (2012), who finds that the largest estimates are generated by cross-sectional data and the smallest by time-series. This is consistent with, for example, findings of Taheri (1994), who uses time-series data to estimate the elasticity for US manufacturing and Halvorsen (1977), who uses panel data and whose estimates are somewhat larger. Because most estimates are produced when using time-series data, this becomes our baseline category. Two reference categories are panel and cross-sectional data. Then, we account for the frequency of data used to generate the estimates. Most researchers use annual data, although some also collected monthly, quarterly, and in a few cases, 5-year data to generate the estimates. Thus, we create a dummy control with a baseline category including studies with annual data. The first reference category consists of studies using high frequency data, or, in other words, quarterly and monthly data. The second reference category is for low frequency data. We also include a variable that captures the year for which the elasticity was estimated. This should allow us to identify a potential trend in the elasticity over time. In addition, another variable in this category is the length of the data – the number of years captured by the data used by researchers to produce the estimates.

Structural variation

Several variables were collected to account for structural variation. First of all, we include the logarithm of GDP per capita measured at the midyear of the data used to generate the estimate of the elasticity. As an example, if a study used US state-level data for the years 1970-1990, we would collect GDP per capita for the US in 1980. If a study uses an international sample and estimates the elasticity for the whole sample, we use the world's GDP per capita for the midyear as an approximate value. Similarly, if a study uses pooled sample consisting of several countries, we compute an average GDP per capita for these countries. This way, it is possible to account for regional variation and level of development. For instance, in the meta-analysis concerning capital-energy substitution, Koetse et al. (2008) find that there are noticeable differences in the elasticity between regions. It is not unlikely that a similar pattern could emerge in the interfuel substitution and thus, this work tries to address this potential source of heterogeneity. Then, we distinguish between the state-level estimates, which we choose to be a baseline category, industrial (including manufacturing as a whole) level estimates, subindustrial (including submanufacturing sectors) level estimates and estimates for the electric power generation sector as they might markedly differ in magnitudes.

Design of the production function

Although translog framework, our baseline category, dominates in the reviewed literature, some researchers chose a different approach. Besides authors of more than 70% of the estimates using the translog function, second most used specification is the linear logit model, which is our first reference category. The linear logit model was estimated by, for instance, Jones (1995), Considine (2018), Urga and Walters (2003) or Steinbuks and Narayanan (2015). Only two studies adopted a framework different from the translog or linear logit model and thus, we merge these studies, creating our second reference category. Moreover, if estimates produced by a different framework represent less than 3% out of all estimates, we merge these studies with those using the linear logit model. A slightly different approach is chosen by Hall (1983); Hall (1986b); Hall (1986a). He tests a number of important restrictions for a translog model and found some violations of strict conditions. Therefore, he estimated non-homothetic translog models instead of homothetic ones. To control for this, we create a dummy equal to one if the estimated model is non-homothetic and zero otherwise.

Estimation technique

Across studies used in this meta-analysis, the prevailing estimation technique is seemingly unrelated regression. More than 60% of all estimates were generated using this technique, followed by approximately 20% of the estimates produced while employing maximum likelihood estimation. As an exception, Lin and Tian (2017) employ three-stage least squares and Ma and Stern (2016) and Considine (2018) estimate their models using the generalized method of moments (GMM). Our baseline category is therefore studies employing seemingly unrelated regression and two reference categories are studies using maximum likelihood estimation and studies utilizing a different estimation technique. We also control for dynamic specifications and include a dummy equal to one if the estimated model is dynamic. As argued by Wang et al. (2019), the static model might not properly reflect the adjustment process of energy price change and thus, some authors estimate the dynamic model. Such approach is chosen by, for example, Urga and Walters (2003) or Cho et al. (2004). Concerning dynamic models, we also introduce a dummy variable for long-run effect estimates. However, most researchers do not explicitly state whether they intend to estimate the long-run or short-run elasticity of demand. Some studies state that their estimates represent long-run or short-run substitution possibilities, but they often rely only on their interpretations rather than on explicit formulations (Uri, 1978; Griffin, 1977). Therefore, we only refer to elasticities as long-run estimates if they were produced by a dynamic model and explicitly specified as long-run effect estimates of the elasticity.

Publication characteristics

Because it is not unlikely that papers published in trusted journals or repeatedly cited studies employ more rigorous and appropriate statistical methods to analyze interfuel substitution, we also control for publication characteristics. Firstly, we use the recursive discounted impact factor from RePEc to reflect study quality. If a study was published in a journal that does not appear in the RePEc list, we set this variable equal to 0. Secondly, we include a variable capturing number of study citations in Google Scholar per year.

We aim to regress all explanatory variables on our dependent variable, the elasticity of demand. However, many of these variables might prove to be unimportant in the regression, possibly leading to over-fitting and low precision of the model. Our goal is to choose the most relevant subset of variables capturing the essential drivers of the true effect, but due to the relatively high number of collected variables, deciding which variables should be included in the final model would be difficult and time-consuming. We face substantial model uncertainty.

If we were to manually select appropriate explanatory variables, we would have to run $2^{26} \approx 67000000$ possible models with all the different combinations. Because such procedure is infeasible, this problem can be addressed by using model averaging techniques, a standard method used in meta-analyses. These techniques estimate all possible 2^{26} specifications and assign weights to them so that better-performing models receive larger weights than the worse ones. Firstly, this work employs the Bayesian model averaging, also used by, for example, Bajzik et al. (2020) or Gechert et al. (2021). Additionally, Frequentist model averaging is used as a robustness check.

Bayesian model averaging (BMA) technique fits a large number of specifications for all possible variable combinations and then, it calculates the weighted average of these regressions. According to Kamenická (2021), we will now define several statistical measures used by BMA:

- *Posterior model probability* (PMP) can be considered an equivalent to information criteria. In BMA, PMP measures the model's performance compared to other models. PMPs are also used as weights to calculate the weighted average of all the regressions and they are calculated based on Bayes' theorem.

- *Posterior inclusion probability* (PIP) is the probability of a variable being included in the model. It is analogous to the statistical significance – the higher the PIP, the more important variable is in explaining the differences across estimates. This work will follow common standards and assume that if $0.5 < PIP < 0.75$, the variable is weakly significant, if $0.75 < PIP < 0.95$ it is substantially significant and if $0.95 < PIP < 0.99$ the variable is considered strongly significant. If PIP is higher than 0.99, it implies decisive significance (Jeffreys, 1998).

- *Weighted posterior mean* (WPM) could be viewed as a model average parameter estimate. It is calculated based on individual model estimates weighted by their PMP.

- *Weighted posterior variance* is calculated based on the weighted average of the models' variances and weighted variance of $\beta_j s$ across all the models.

- *Weighted posterior standard deviation* (WPSD) is then analogous to a standard error of the estimate.

Because the implementation of the BMA is quite demanding, this work employs the Markov chain Monte Carlo algorithm (MCMC). In order to make application of BMA practically possible, this method focuses on the most convenient models with the highest PIPs.

BMA also requires us to formulate some assumptions (Hoeting et al., 1999). For the estimation, we use the Bayesian model selection package for R software, developed by Zeugner and Feldkircher (2015), and we choose the following baseline arguments:

- $burn = 100000$
  $Burn$ denotes the number of iterations that are not stored, also called $burn - ins$. We set this argument to 100 000.

- $iter = 300000$
  $iter$ stands for the total number of iterations to be sampled. In our case, we choose 300 000 iterations.

- $g = "UIP"$
  For the g-prior g, which indicates how much weight we prefer to give to our prior on regression coefficients, we choose the most popular one, called unit information prior (UIP). This way, the prior is assigned weight proportional to weight of one observation in the dataset.

- $mprior = "dilut"$
  As the prior model probability, we use dilution prior due to George (2010), which prevents multicollinearity. Because this prior compensates for redundancy between model classes, it is superior to other priors and it is our preferred prior.

- $nmodel = 50000$
  $nmodel$ denotes the number of best models for which we store the information. In our case, this argument is set to 50 000.

- $mcmc = "bd"$
  $mcmc$ stands for the Markov chain Monte Carlo algorithm, used to calculate posterior probabilities. Within this algorithm, there are multiple ways, also called samplers, how to choose candidate model in every iteration. Our choice is the birth-death or bd sampler. This sampler randomly draws one explanatory variable and then, it chooses to drop or add it based on whether this particular variable is included in the previous model or not.

5.2 Results and robustness check

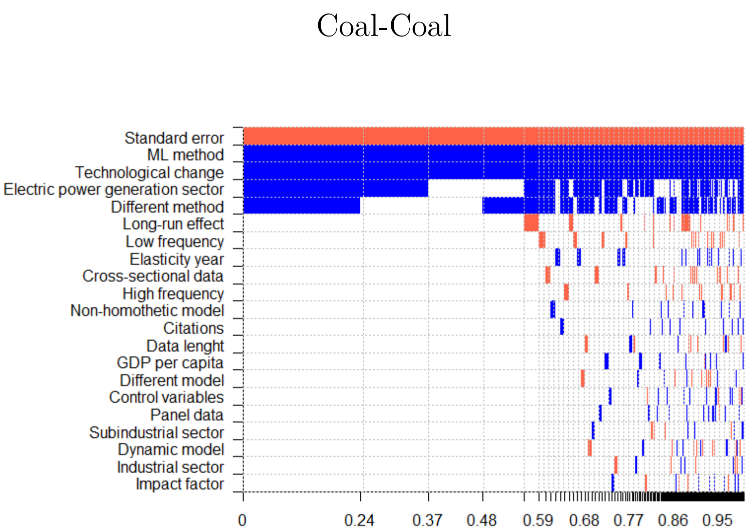
In this section, we present the results from the model averaging estimation. In particular, we will focus on the results from the BMA estimation, as this framework is superior. FMA estimation serves as a robustness check. For the sake of transparency, we will firstly discuss the results for the own-price elasticity of demand for coal. Then, we will jointly discuss the cross-price elasticity of demand between coal and other fuels in both directions. To wrap up, we summarize the main findings and pinpoint the most important factors driving the estimates.

Own-price elasticity for coal

Figure 5.1 reports the results from the BMA estimation, performed on the data winsorized at 1% level, for the own-price elasticity for coal. The horizontal axis presents the posterior model probabilities sorted from the lowest to the highest so that the best performing models are on the left. The vertical axis depicts the variables sorted by their relevance so that the variables with the highest PIP are at the top. Blue colour (darker in grayscale) indicates that the variable has a positive effect on the elasticity, whereas red colour (lighter in grayscale) implies that the variable's estimated coefficient is negative. White colour means that the variable is excluded from the corresponding regression model. Table 5.2 reports numeric results from the BMA estimation, as well as results from the alternative model averaging technique called FMA as a robustness check. The results of BMA using alternative priors are shown in Figure C.1 in Appendix C.

The standard error is significant in the regression, indicating that publication bias survives even after controlling for collected explanatory variables. Regarding the other drivers of heterogeneity, employing estimation techniques other than SUR and capturing technological change is linked to estimates smaller in the absolute value, as the estimated coefficients are positive. Besides, the own-price elasticity for coal is found to be significantly smaller in absolute value in the electric power generation sector than the elasticity estimated at the macro-level. The remaining variables, including data dimension, variables capturing other structural variation, the design of the production function, or publication characteristics do not seem to have a significant impact on the own-price elasticity.

Figure 5.1: Model inclusion in Bayesian model averaging, own-price elasticity for coal



Notes: The figure depicts the results of BMA estimation. The response variable is the estimate of the own-price elasticity of demand for coal. Each column represents a model; the horizontal axis depicts the posterior model probabilities sorted from the lowest to the highest so that the best models are on the left-side; the vertical axis depicts the explanatory variables sorted by their relevance so that variables with highest PIP are at the top. Blue colour (darker in grayscale) = the variable has a positive effect on the elasticity; red colour (lighter in grayscale) = the variable has a negative effect on the elasticity; white colour = the variable is excluded from the model. Numerical results of the BMA estimation are presented in Table 5.2. GDP = gross domestic product, ML = Maximum Likelihood.

Table 5.2: Explaining the heterogeneity in the own-price elasticity for coal

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Intercept	**1.00**	-1.12		-1.30	0.96	**0.17**
Standard error	**1.00**	-0.48	0.07	-0.49	0.08	**0.00**
Specifications						
Control variables	0.03	0.00	0.03	0.00	0.41	0.00
Technological change	**1.00**	0.55	0.13	0.56	0.21	**0.01**
Data characteristics						
Panel data	0.03	0.00	0.03	-0.00	0.04	0.00
Cross-sectional data	0.05	-0.02	0.09	-0.21	0.93	0.82
High frequency	0.03	-0.01	0.07	-0.13	0.57	0.82
Low frequency	0.06	-0.02	0.08	-0.15	0.23	0.52
Elasticity year	0.05	0.00	0.00	0.00	0.01	0.91
Data lenght	0.03	-0.00	0.00	-0.00	0.01	0.76
Structural variation						
GDP per capita	0.03	0.00	0.01	0.02	0.09	0.78
Industrial sector	0.03	0.00	0.02	0.00	0.05	0.00
Subindustrial sector	0.03	-0.00	0.03	0.00	0.26	0.00
Electric power generation sector	**0.68**	0.42	0.33	0.54	0.25	**0.03**
Design of the production function						
Different model	0.03	-0.00	0.04	-0.00	0.31	0.00
Non-homothetic model	0.03	0.01	0.08	0.08	0.30	0.80
Estimation technique						
ML method	**1.00**	0.84	0.16	0.78	0.23	**0.00**
Different method	**0.61**	0.44	0.41	0.78	0.44	**0.08**
Dynamic model	0.03	-0.00	0.03	0.00	0.03	0.00
Long-run effect	0.09	-0.03	0.12	-0.34	0.29	0.24
Publication characteristics						
Impact factor	0.02	0.00	0.02	0.00	0.26	0.00
Citations	0.03	0.00	0.01	0.03	0.05	0.59
Studies	42			42		
Observations	193			193		

Notes: The table shows numerical results of the BMA estimation on the left-side and FMA estimation on the right-side. The response variable is the estimate of the own-price elasticity for coal. PIP = posterior inclusion probability, PM = posterior mean, PSD = posterior standard deviation, Coef. = estimated coefficient, SE = standard error, GDP = gross domestic product, ML= maximum likelihood.

Coal-Electricity substitution

Figure 5.2 and Table 5.3 report results from the BMA and FMA estimation for the coal-electricity and pairwise electricity-coal elasticity. The results of BMA using alternative priors are shown in Figure C.2 in Appendix C.

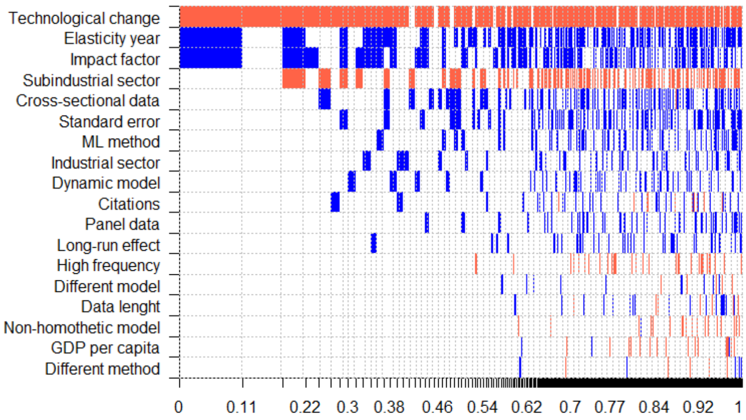
The standard error does not have significant impact on the estimates of the coal-electricity elasticity. On the other hand, standard error is significant explanatory variable in the case of the electricity-coal elasticity. Thus, these results provide support for the previously used tests to detect publication bias. The estimates of the electricity-coal elasticity are found to be systematically affected by selective reporting even after controlling for other sources of heterogeneity.

An important variable that explains the heterogeneity in the estimates of the cross-price elasticity of demand between coal and electricity is technological change. It has a negative effect on the estimates. Moreover, the variable capturing the elasticity year is deemed weakly significant and has a positive estimated coefficient. This implies that the estimates of the elasticity increases over time. The last weakly significant variable based on the BMA estimation is the impact factor with a positive effect on the estimates. This indicates that studies published in journals with higher impact factor are linked to higher estimates of the cross-price elasticity between coal and electricity.

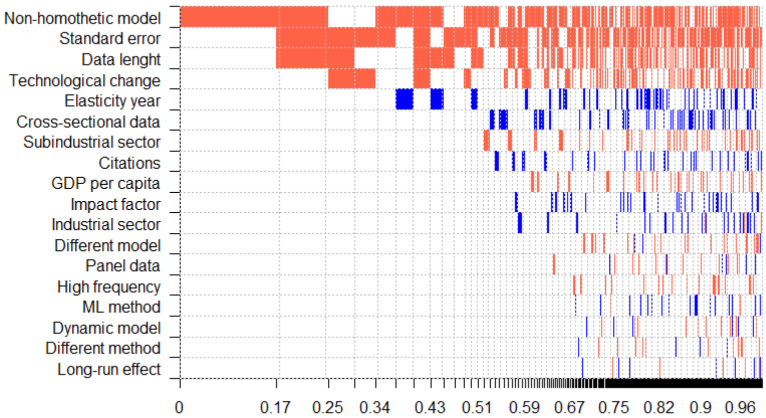
On the other hand, apart from the standard error, the only variable significantly impacting the estimates of the electricity-coal elasticity is the variable capturing studies that estimated a non-homothetic model instead of a homothetic one. Estimation of such model is associated with smaller estimates of the elasticity between electricity and coal.

Figure 5.2: Model inclusion in Bayesian model averaging, coal-electricity substitution

(a) Coal-Electricity



(b) Electricity-Coal



Notes: The figure depicts the results of BMA estimation. The response variables are the estimate of the cross-price elasticity of demand between coal and electricity at the top and estimate of the cross-price elasticity of demand between electricity and coal at the bottom. Each column represents a model; the horizontal axis depicts the posterior model probabilities sorted from the lowest to the highest so that the best models are on the left-side; the vertical axis depicts the explanatory variables sorted by their relevance so that variables with highest PIP are at the top. Blue colour (darker in grayscale) = the variable has a positive effect on the elasticity; red colour (lighter in grayscale) = the variable has a negative effect on the elasticity; white colour = the variable is excluded from the model. Numerical results of the BMA estimation are presented in Table 5.3. GDP = gross domestic product, ML = Maximum Likelihood.

Table 5.3: Explaining the heterogeneity in the coal-electricity substitution

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel A: Coal-Electricity						
Intercept	**1.00**	-0.05		-1.05	0.61	**0.08**
Standard error	0.22	0.09	0.19	0.38	0.22	0.07
Specifications						
Technological change	**0.93**	-0.55	0.22	-0.54	0.17	**0.00**
Data characteristics						
Panel data	0.09	0.03	0.10	0.37	0.34	0.28
Cross-sectional data	0.28	0.31	0.56	1.10	0.76	0.15
High frequency	0.05	-0.02	0.13	-0.06	0.28	0.83
Elasticity year	**0.55**	0.01	0.01	0.03	0.01	**0.01**
Data lenght	0.04	0.00	0.00	0.01	0.02	0.61
Structural variation						
GDP per capita	0.03	-0.00	0.02	-0.00	0.06	1.00
Industrial sector	0.13	0.04	0.13	-0.03	0.17	0.88
Subindustrial sector	0.35	-0.20	0.30	-0.55	0.28	0.05
Design of the production function						
Different model	0.04	0.00	0.06	-0.26	0.38	0.49
Non-homothetic model	0.03	-0.00	0.06	-0.00	0.23	1.00
Estimation technique						
ML method	0.14	0.06	0.17	0.59	0.33	0.07
Different method	0.03	0.00	0.05	-0.00	0.28	1.00
Dynamic model	0.13	0.04	0.13	0.20	0.26	0.42
Long-run effect	0.08	0.03	0.11	0.07	0.20	0.73
Publication characteristics						
Impact factor	**0.53**	0.40	0.42	0.52	0.41	**0.20**
Citations	0.10	0.01	0.04	-0.09	0.13	0.48
Studies	30			30		
Observations	112			112		

Continued on next page

Table 5.3: Explaining the heterogeneity in the coal-electricity substitution (continued)

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel B: Electricity-Coal						
Intercept	**1.00**	0.20		0.59	0.32	**0.06**
Standard error	**0.63**	-0.97	0.86	-1.47	0.48	**0.00**
Specifications						
Technological change	0.31	-0.04	0.07	-0.11	0.05	0.04
Data characteristics						
Panel data	0.03	-0.00	0.01	-0.05	0.07	0.48
Cross-sectional data	0.11	0.03	0.08	0.23	0.18	0.21
High frequency	0.03	-0.00	0.03	-0.22	0.15	0.15
Elasticity year	0.17	0.00	0.00	0.00	0.00	0.21
Data lenght	0.46	-0.00	0.00	-0.01	0.00	0.06
Structural variation						
GDP per capita	0.07	-0.00	0.01	-0.04	0.03	0.14
Industrial sector	0.05	0.00	0.02	0.09	0.08	0.23
Subindustrial sector	0.08	-0.01	0.04	-0.09	0.08	0.28
Design of the production function						
Different model	0.04	-0.00	0.02	-0.13	0.10	0.20
Non-homothetic model	**0.75**	-0.29	0.20	-0.30	0.12	**0.01**
Estimation technique						
ML method	0.03	0.00	0.01	0.08	0.08	0.31
Different method	0.03	-0.00	0.02	0.00	0.02	0.00
Dynamic model	0.03	-0.00	0.01	0.08	0.08	0.34
Long-run effect	0.02	0.00	0.01	0.00	0.01	0.00
Publication characteristics						
Impact factor	0.06	0.01	0.03	-0.03	0.12	0.83
Citations	0.07	0.00	0.01	0.01	0.03	0.75
Studies	31			31		
Observations	114			114		

Notes: The table shows numerical results of the BMA estimation on the left-side and FMA estimation on the right-side. The response variables are the estimate of the cross-price elasticity between coal and electricity (*Panel A*) and estimate of the cross-price elasticity of demand between electricity and coal (*Panel B*). PIP = posterior inclusion probability, PM = posterior mean, PSD = posterior standard deviation, Coef. = estimated coefficient, SE = standard error, GDP = gross domestic product, ML= maximum likelihood.

Coal-Gas substitution

Results of the BMA and FMA estimation for the coal-gas and gas-coal elasticity are presented in Figure 5.3 and Table 5.4. In the case of gas-coal elasticity, we use data winsorized at 1% level. The results of BMA using alternative priors are shown in Figure C.3 in Appendix C.

Publication bias is not significantly impacting the estimates of the coal-gas elasticity based on the BMA estimation. This is in line with the estimates from publication bias tests, which detected either insignificant or little selective reporting. However, the tests detected publication bias in the estimates of the gas-coal elasticity. Even though our tests systematically suggested that publication bias is present, the original BMA estimation did not find standard error to be a significant explanatory variable. Because the results were not consistent, we decided to examine this. In order to identify whether there exists a specific subset of studies driving the bias, we tried adding several interaction terms and re-running the BMA estimation. We came to a conclusion that the bias is most likely to be driven by the industry-level estimates.[1] The interaction term *SE Industrial sector* is highly significant in the regression and we report results from this estimation.[2] Because the authors might often assume that industrial-level estimates should be relatively large and positive (Stern, 2012), this does not seem unreasonable. The researchers might be tempted to hunt for estimates that are greater in value to provide results in line with their intuition and the existing research.

Regarding other drivers of the heterogeneity, the first important explanatory variable is the estimation technique. The results indicate that methods differ-

[1] To provide more evidence, we performed additional linear tests to detect publication bias. We distinguished between two panels – the first panel only included industry-level estimates, while the second panel excluded industry-level estimates. The panels consisted of 49 and 55 estimates, respectively. The results from regressions support our choice. Very severe publication bias is detected in the panel including only the industry-level estimates, and the mean beyond bias is not significantly different from zero. On the contrary, there is no publication bias detected in the second panel. The results are shown in Table C.1 in Appendix C.

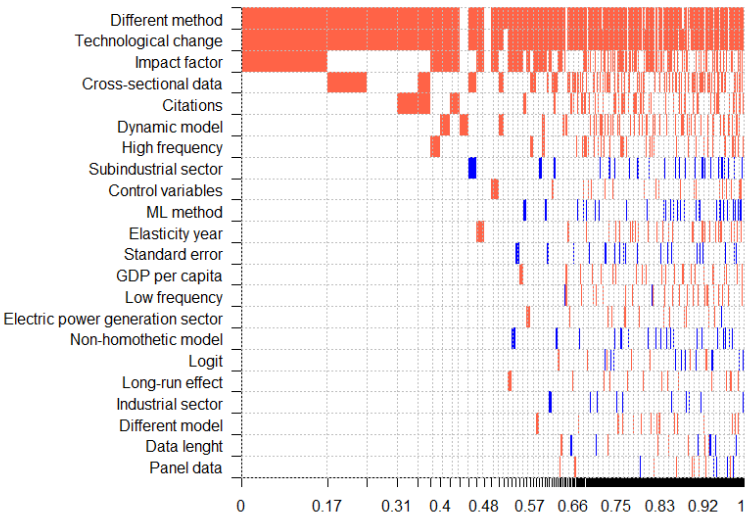
[2] We also examined whether industry-level estimates could drive publication bias in other subsets. However, this interaction term is not significant in any other subset. Publication bias seems to be driven by the industry-level estimates only in the case of gas-coal elasticity. In Table 4.1, we showed that rather than being a systematic problem in the literature, it seems that certain studies cause publication bias. Thus, the fact that the industry-level estimates drive publication bias only in the case of gas-coal elasticity is perhaps accidental and caused by a few reviewed papers.

ent from SUR and ML estimation produce systematically smaller estimates of the cross-price elasticity of demand between coal and gas, as suggested by the negative coefficient. In addition, modeling technological change is also linked to smaller estimates of the coal-gas elasticity.

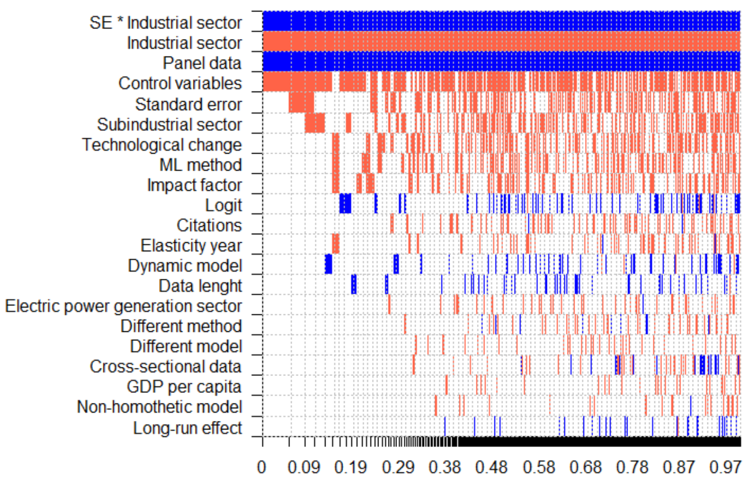
On the other hand, apart from the standard error, control variables other than those capturing technological change seem to matter in the case of pairwise elasticity between gas and coal. The estimated coefficient is negative, implying that omitting other endogeneity controls perhaps leads to exaggerated estimates of the elasticity. Another important variable explaining the heterogeneity among estimates is data dimension, as it seems that panel data produce larger estimates than time-series data, our baseline category. Lastly, after controlling for publication bias, the industry-level estimates are estimated to be significantly smaller than the state-level estimates. This provides further backing to our notion that the publication bias is driven by the industry-level estimates. Since they are estimated to be smaller than the state-level estimates, researchers perhaps try to hunt for large estimates and publication bias arises.

Figure 5.3: Model inclusion in Bayesian model averaging, coal-gas substitution

(a) Coal-Gas



(b) Gas-Coal



Notes: The figure depicts the results of BMA estimation. The response variables are the estimate of the cross-price elasticity of demand between coal and gas at the top and estimate of the cross-price elasticity of demand between gas and coal at the bottom. Each column represents a model; the horizontal axis depicts the posterior model probabilities sorted from the lowest to the highest so that the best models are on the left-side; the vertical axis depicts the explanatory variables sorted by their relevance so that variables with highest PIP are at the top. Blue colour (darker in grayscale) = the variable has a positive e ect on the elasticity; red colour (lighter in grayscale) = the variable has a negative e ect on the elasticity; white colour = the variable is excluded from the model. Numerical results of the BMA estimation are presented in Table 5.3. GDP = gross domestic product, ML = Maximum Likelihood.

Table 5.4: Explaining the heterogeneity in the coal-gas substitution

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel A: Coal-Gas						
Intercept	**1.00**	0.73		1.95	1.60	**0.22**
Standard error	0.05	0.01	0.03	0.07	0.09	0.43
Specifications						
Control variables	0.06	-0.01	0.07	-0.29	0.30	0.33
Technological change	**0.92**	-0.48	0.19	-0.48	0.17	**0.01**
Data characteristics						
Panel data	0.02	-0.00	0.02	-0.00	0.12	0.00
Cross-sectional data	0.29	-0.20	0.35	-0.48	0.41	0.25
High frequency	0.10	-0.05	0.18	-0.33	0.40	0.42
Low frequency	0.04	-0.01	0.08	-0.19	0.25	0.44
Elasticity year	0.05	-0.00	0.00	-0.00	0.01	0.64
Data lenght	0.03	0.00	0.00	-0.00	0.01	0.00
Structural variation						
GDP per capita	0.05	-0.01	0.04	-0.12	0.15	0.41
Industrial sector	0.03	0.00	0.03	0.00	0.08	0.00
Subindustrial sector	0.07	0.02	0.10	0.13	0.24	0.57
Electric power generation sector	0.04	-0.01	0.05	0.02	0.23	0.92
Design of the production function						
Logit	0.04	0.00	0.06	0.49	0.38	0.19
Different model	0.03	-0.01	0.07	-0.00	0.15	0.00
Non-homothetic model	0.04	0.01	0.06	0.09	0.20	0.63
Estimation technique						
ML method	0.06	0.01	0.06	0.21	0.22	0.33
Different method	**0.93**	-1.05	0.37	-1.25	0.41	**0.00**
Dynamic model	0.14	-0.04	0.13	-0.27	0.23	0.22
Long-run effect	0.03	-0.01	0.05	-0.00	0.05	0.00
Publication characteristics						
Impact factor	0.49	-0.18	0.21	-0.21	0.22	0.32
Citations	0.19	-0.02	0.04	-0.03	0.06	0.64
Studies	25			25		
Observations	124			124		

Continued on next page

Table 5.4: Explaining the heterogeneity in the coal-gas substitution (continued)

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel B: Gas-Coal						
Intercept	**1.00**	0.78		0.99	0.71	**0.16**
Standard error	0.38	-0.02	0.03	-0.04	0.02	0.14
SE * Industrial sector	**1.00**	4.95	0.85	5.00	0.91	**0.00**
Specifications						
Control variables	**0.77**	-0.56	0.37	-0.67	0.34	**0.04**
Technological change	0.35	-0.14	0.22	-0.34	0.22	0.13
Data characteristics						
Panel data	**1.00**	0.77	0.19	0.75	0.22	**0.00**
Cross-sectional data	0.08	-0.01	0.16	0.55	0.61	0.37
Elasticity year	0.14	-0.00	0.00	-0.01	0.01	0.49
Data lenght	0.12	0.00	0.00	0.01	0.01	0.50
Structural variation						
GDP per capita	0.07	-0.01	0.04	0.00	0.06	0.00
Industrial sector	**1.00**	-0.97	0.22	-0.98	0.26	**0.00**
Subindustrial sector	0.38	-0.19	0.28	-0.51	0.30	0.09
Electric power generation sector	0.11	-0.03	0.13	0.10	0.29	0.73
Design of the production function						
Logit	0.17	0.07	0.21	0.27	0.34	0.43
Different model	0.10	-0.04	0.17	-0.34	0.48	0.48
Non-homothetic model	0.06	-0.02	0.10	-0.00	0.17	0.00
Estimation technique						
ML method	0.30	-0.13	0.24	-0.13	0.23	0.57
Different method	0.10	-0.05	0.21	-0.63	0.63	0.32
Dynamic model	0.12	0.04	0.12	0.02	0.18	0.92
Long-run effect	0.05	0.01	0.05	0.00	0.01	0.00

Continued on next page

Table 5.4: Explaining the heterogeneity in the coal-gas substitution (continued)

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Publication characteristics						
Impact factor	0.27	-0.10	0.19	-0.45	0.38	0.23
Citations	0.15	-0.02	0.05	-0.01	0.09	0.94
Studies	25			25		
Observations	104			104		

Notes: The table shows numerical results of the BMA estimation on the left-side and FMA estimation on the right-side. The response variables are the estimate of the cross-price elasticity between coal and gas (*Panel A*) and estimate of the cross-price elasticity of demand between gas and coal (*Panel B*). PIP = posterior inclusion probability, PM = posterior mean, PSD = posterior standard deviation, Coef. = estimated coe□cient, SE = standard error, GDP = gross domestic product, ML= maximum likelihood.

Coal-Oil substitution

The results for the Coal-Oil and Oil-Coal elasticity are presented graphically in Figure 5.4 and numerically in Table 5.5. For the estimation, we used data winsorized at 1% level. The results of BMA using alternative priors are shown in Figure C.4 in Appendix C.

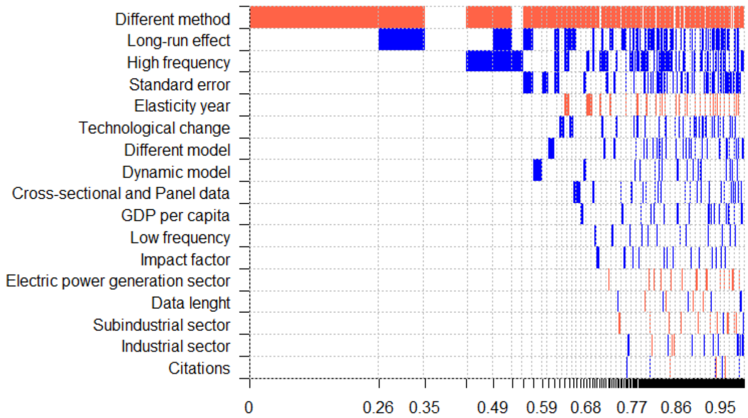
Based on the BMA specification, the standard error is not an important explanatory variable in the case of coal-oil elasticity. On the contrary, it is significant in the case of cross-price elasticity between oil and coal. The results therefore indicate that the estimates of the oil-coal elasticity are affected by selective reporting even after controlling for other potential sources of heterogeneity.

The only variable affecting the coal-oil elasticity is using estimation techniques other than SUR. The estimated coefficient is negative. No other variable has a significant impact on these estimates according to the BMA estimation.

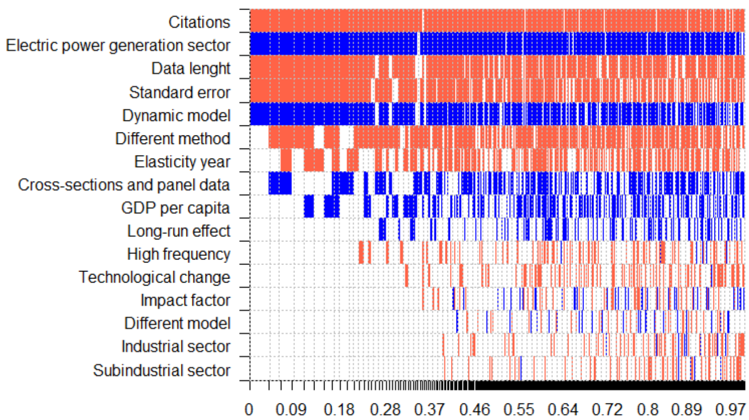
In contrast, there are several significant variables explaining the heterogeneity in the estimates of the elasticity between oil and coal. Firstly, the estimated, significant effect of the variable capturing number of citations per year is negative, indicating that more cited studies tend to report smaller estimates of the elasticity between oil and coal. Employing estimation methods different from SUR and using data over longer period of time has a negative impact on the final estimates. According to Bacon and Mundial (1992), it could be that using too lengthy period of adjustment leads to underestimated price responsiveness. On the contrary, estimating a dynamic model is linked to larger estimates. As stated in Wang et al. (2019), the static model may not properly reflect the adjustment process of energy price change. Thus, the dynamic model might perform better and allow for greater estimates of the elasticity. Interestingly, the elasticity is found to be significantly greater in the electric power generation sector than at the state-level, our baseline category. In addition, using cross-sectional or panel data produces greater estimates of the oil-coal elasticity than using time-series, although the coefficient is only marginally significant based on the BMA. The last marginally significant variable is elasticity year. As implied by the negative coefficient, oil-coal elasticity follows a negative trend. In other words, the estimates of the elasticity seem to slowly decrease over time.

Figure 5.4: Model inclusion in Bayesian model averaging, coal-oil substitution

(a) Coal-Oil



(b) Oil-Coal



Notes: The figure depicts the results of BMA estimation. The response variables are the estimate of the cross-price elasticity of demand between coal and oil at the top and estimate of the cross-price elasticity of demand between oil and coal at the bottom. Each column represents a model; the horizontal axis depicts the posterior model probabilities sorted from the lowest to the highest so that the best models are on the left-side; the vertical axis depicts the explanatory variables sorted by their relevance so that variables with highest PIP are at the top. Blue colour (darker in grayscale) = the variable has a positive effect on the elasticity; red colour (lighter in grayscale) = the variable has a negative effect on the elasticity; white colour = the variable is excluded from the model. Numerical results of the BMA estimation are presented in Table 5.3. GDP = gross domestic product.

Table 5.5: Explaining the heterogeneity in the coal-oil substitution

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel A: Coal-Oil						
Intercept	**1.00**	0.50		0.41	0.61	**0.51**
Standard error	0.15	0.02	0.05	0.16	0.07	0.03
Specifications						
Technological change	0.07	0.01	0.06	0.21	0.13	0.13
Data characteristics						
Cross-sectional and Panel data	0.06	0.01	0.06	0.02	0.08	0.79
High frequency	0.30	0.20	0.34	0.71	0.29	0.01
Low frequency	0.04	0.01	0.07	0.04	0.17	0.80
Elasticity year	0.09	-0.00	0.00	-0.01	0.01	0.06
Data lenght	0.02	0.00	0.00	-0.00	0.01	0.88
Structural variation						
GDP per capita	0.04	0.00	0.02	0.03	0.08	0.70
Industrial sector	0.02	0.00	0.02	0.00	0.15	0.00
Subindustrial sector	0.02	-0.00	0.02	0.00	0.04	0.00
Electric power generation sector	0.03	-0.01	0.05	-0.13	0.26	0.60
Design of the production function						
Different model	0.07	0.03	0.13	0.51	0.31	0.10
Estimation technique						
Different method	**0.74**	-0.38	0.27	-0.72	0.21	**0.00**
Dynamic model	0.06	0.02	0.09	0.03	0.16	0.86
Long-run effect	0.33	0.24	0.38	0.62	0.31	0.04
Publication characteristics						
Impact factor	0.03	0.00	0.04	0.01	0.13	0.94
Citations	0.02	0.00	0.01	0.00	0.04	0.00
Studies	26			26		
Observations	135			135		

Continued on next page

Table 5.5: Explaining the heterogeneity in the coal- oil substitution (continued)

Response variable: Estimated elasticity	Bayesian model averaging (baseline model)			Frequentist model averaging (robustness check)		
	PIP	PM	PSD	Coef.	SE	p-value
Panel B: Oil-Coal						
Intercept	**1.00**	0.40		0.05	0.36	**0.90**
Standard error	**0.82**	-0.04	0.03	-0.05	0.02	**0.01**
Specifications						
Technological change	0.17	-0.02	0.06	-0.12	0.10	0.23
Data characteristics						
Cross-sectional and panel data	**0.52**	0.13	0.16	0.12	0.11	**0.28**
Elasticity year	**0.53**	-0.00	0.00	-0.01	0.00	**0.02**
Data lenght	**0.84**	-0.02	0.01	-0.02	0.01	**0.00**
Structural variation						
GDP per capita	0.43	0.03	0.05	0.10	0.04	0.03
Industrial sector	0.11	-0.01	0.04	0.00	0.00	0.00
Subindustrial sector	0.11	-0.00	0.03	0.00	0.01	0.00
Electric power generation sector	**0.96**	0.50	0.18	0.60	0.14	**0.00**
Design of the production function						
Diﬀerent model	0.12	-0.00	0.07	-0.00	0.03	0.00
Estimation technique						
Diﬀerent method	**0.75**	-0.21	0.15	-0.33	0.11	**0.00**
Dynamic model	**0.82**	0.34	0.21	0.45	0.16	**0.00**
Long-run eﬀect	0.20	0.05	0.12	0.13	0.17	0.44
Publication characteristics						
Impact factor	0.16	-0.01	0.07	-0.16	0.12	0.20
Citations	**0.97**	-0.14	0.05	-0.09	0.05	**0.06**
Studies	24			24		
Observations	111			111		

Notes: The table shows numerical results of the BMA estimation on the left-side and FMA estimation on the right-side. The response variables are the estimate of the cross-price elasticity between coal and oil (*Panel A*) and estimate of the cross-price elasticity of demand between oil and coal (*Panel B*). PIP = posterior inclusion probability, PM = posterior mean, PSD = posterior standard deviation, Coef. = estimated coeﬃcient, SE = standard error, GDP = gross domestic product.

To sum up, there are several aspects driving the results. The publication bias survives after controlling for all the other aspects in four cases. The standard error has a significant impact on the estimates of the own-price elasticity for coal, electricity-coal, gas-coal, and oil-coal elasticity. In the remaining cases, the estimates seem to be driven by different features. One might wonder why publication bias affects only the own-price elasticity and the cross-price elasticity between other fuels and coal. Why the cross-price elasticity between coal and alternative fuels remains largely unaffected? In the case of own-price elasticity for coal, publication bias is not unexpected. Because economic theory does not provide much support for the positive own-price elasticity, the researchers tend to under-report clearly incorrect estimates. Publication bias plagues the literature. This issue was already recognized by Stern (2012). However, the reasons why the cross-price elasticity between other fuels and coal suffers from substantial publication bias while the elasticity in the opposite direction does not are not so clear. Nonetheless, the fact that the means beyond bias are consistently estimated to be zero provides some reasoning. The cross-price elasticity might be intuitively assumed to be significantly different from zero, but the true effects are in fact estimated to be very close to zero. Thus, researchers could be tempted to selectively report greater estimates or tweak the estimation procedure so that the generated estimates are larger. If the true cross-price elasticity between coal and other fuels is greater than 0, there is no need for such selective reporting. On the contrary, if the cross-price elasticity between other fuels and coal is not significantly different from zero, substantial publication bias arises.

Apart from the standard errors, one of the most important factors seems to be endogeneity controls included in the model used to generate the estimates. Using control variables other than those capturing technological change is linked to smaller estimates of the elasticity between gas and coal. The problem of potential omitted variable bias is extensively discussed in, for example, Pettersson et al. (2012). Bias might stem from several different sources, such as omitting the role of nuclear energy and hydropower or neglecting the impact of government policies. Similarly, in the case of own-price elasticity for coal, coal-electricity, and coal-gas elasticity, modeling technological change leads to estimates that are smaller in absolute value. This is in line with the argumentation provided by Stern (2012), who believes it to be the most important omitted variable. Following Welsch and Ochsen (2005), the importance of technological change is also recognized in Ma et al. (2008), Urga and Walters (2003), or Lin and Tian (2017). However, we note that modeling technological change is not trivial and if not modeled correctly, it could lead to twisted estimates, as shown in Borjas et al. (2012). This has to be kept in mind when assessing the impact of technological change on the estimates.

Regarding the different settings of studies, including data characteristics, design of production functions, or estimation techniques, mainly estimation methods other than SUR systematically impact the estimated elasticities. It is linked to the estimates of the own-price elasticity for coal, coal-gas, coal-oil, and oil-coal elasticity that are smaller in absolute value. Unlike Stern (2012) and Bacon and Mundial (1992), we do not find any persisting and systematic relationship between the values of the estimates and data dimension. This also contradicts findings of Griffin (1977) or Taheri (1994). They compare their results across studies using different data dimensions and suggest that cross-sectional, panel and time-series data could yield estimates that systematically differ in magnitudes. We found that only estimates of the gas-coal and oil-coal elasticity produced while using panel or cross-sectional data are significantly greater than estimates generated by time-series.

In contrast to Stern (2012) or Shahiduzzaman and Alam (2014), who find that estimates are greater at the lower levels of aggregation, we do not detect any significant differences between macro-level elasticities and elasticities estimated for the industrial or subindustrial sectors. In one case, the industrial-level estimates of the gas-coal elasticity are actually significantly smaller than the state-level estimates. However, we find that the own-price elasticity for coal is smaller in the absolute value in the electric power generation sector. For example, the estimates of the own-price elasticity for coal of Kim (2019), who focuses on the Korean electricity generation sector, are much smaller in absolute value than the results of Shin (1981) or Cho et al. (2004), who estimate the state-level elasticities for Korea. We also find that the cross-price elasticity of demand between oil and coal is greater in the electric power generation sector than at the macro-level. This is in line with findings of Shahiduzzaman and Alam (2014).

Moreover, we lack to find consistent evidence that the level of development is important in explaining the differences across estimates of the elasticity.

The non-existence of such evidence, however, is evidence in itself, perhaps indicating that possibilities to substitute between the four energy inputs are not significantly different for emerging and developed countries. This is in line with findings of Stern (2012), who also fails to discover that the level of development significantly impacts substitution possibilities.

Chapter 6

Conclusion

This work focuses on the topic of fuel substitution, dedicating special attention to coal. We synthesize 893 estimates of the own-price elasticity of demand for coal and pairwise, asymmetric cross-price elasticities of demand between coal and three alternative fuels – electricity, gas, and oil. The estimates are collected from 43 studies. This work presents the first meta-analysis of the own-price elasticity for coal, which is our main contribution. In addition, although meta-analyses on the topic of interfuel substitution exist (Stern, 2012, Chen, 2017), they fail to use rigorous and sophisticated tests to detect publication bias. Moreover, they do not treat for the model uncertainty in their analysis of the variation in the estimates. Thus, we also contribute by addressing both these issues using modern, up-to-date methods.

We find that the own-price elasticity for coal and cross-price elasticities between other fuels and coal suffer from significant publication bias. These reported estimates are substantially exaggerated. The mean corrected own-price elasticity of demand for coal is approximately -0.40 – two times smaller than the mean reported estimate. The mean corrected cross-price elasticities between other fuels and coal are almost zero, suggesting that such substitution is close to impossible. In contrast, we do not find substantial publication bias in the case of cross-price elasticities between coal and alternative fuels. Based on this, we show that some substitution possibilities exist, although they are limited. Next to publication bias, we examine the main causes of heterogeneity in the estimates using 26 explanatory variables capturing several study characteristics. To address model uncertainty, we employ Bayesian model averaging (BMA) as a baseline method and Frequentist model averaging (FMA) as a robustness check. We show that previously detected publication bias survives after controlling for additional 26 variables. Another important factor driving the estimates is whether the study controlled for technological change. Capturing technological change in the model is linked to the estimates of the own-price elasticity for coal, coal-electricity, and coal-gas elasticity that are smaller in absolute value. We also find that employing estimation techniques different from seemingly unrelated regression is associated with smaller estimates of the own-price elasticity for coal and coal-gas, coal-oil, and oil-coal cross-price elasticity. Contrary to Stern (2012) and Bacon and Mundial (1992), we do not find any systematic relationship between the data dimension and the estimates of the elasticity. Neither do we find that the elasticities in the industrial or subindustrial sectors are systematically greater than the elasticities at the state-level. Additionally, we do not detect any significant relationship between the level of development, measured by GDP per capita, and the estimates of the elasticity. Thus, similarly to Stern (2012), we fail to discover that the substitution possibilities are significantly different for developed and developing countries.

Few possible drawbacks of our analysis are in order. The tests used to detect publication bias might suffer from endogeneity. However, we demonstrate that we try to correct for this potential issue. We present the results from the instrumental variable regression and disclose our attempt to use caliper tests. Moreover, our results are clearly conditional on the quality of reviewed studies. If the studies share common misspecification, we are unable to control for it and our analysis could be biased. Thus, our judgement is purely based on the reported empirical research.

Bibliography

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109* (8), 2766–94.

Andrikopoulos, A. A., Brox, J. A., & Paraskevopoulos, C. C. (1989). Interfuel and interfactor substitution in ontario manufacturing, 1962-82. *Applied Economics*, *21* (12), 1667–1681.

Antimiani, A., Costantini, V., & Paglialunga, E. (2015). The sensitivity of climate-economy cge models to energy-related elasticity parameters: Implications for climate policy design. *Economic Modelling*, *51*, 38–52.

Apostolakis, B. E. (1990). Interfuel and energy-capital complementarity in manufacturing industries. *Applied Energy*, *35* (2), 83–107.

Atkinson, S. E., & Halvorsen, R. (1976). Interfuel substitution in steam electric power generation. *Journal of Political Economy*, *84* (5), 959–978.

Bacon, R., & Mundial, B. (1992). *Measuring the possibilities of inter-fuel substitution*. Citeseer.

Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, *127*, 103383.

Begg, C. B., & Berlin, J. A. (1988). Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *151* (3), 419–445.

Bello, M. O., Solarin, S. A., & Yen, Y. Y. (2020). Interfuel substitution, hydroelectricity consumption and co2 emissions mitigation in malaysia: Evidence from a transcendental logarithm (trans-log) cost function framework. *Environmental Science and Pollution Research*, *27* (14), 17162–17174.

Blackorby, C., & Russell, R. R. (1989). Will the real elasticity of substitution please stand up?(a comparison of the allen/uzawa and morishima elasticities). *The American economic review*, *79* (4), 882–888.

Borges, A. M., & Pereira, A. M. (1992). Energy demand in portuguese manufacturing: A two-stage model. *Energy*, *17*(1), 61–77.

Borjas, G. J., Grogger, J., & Hanson, G. H. (2012). Comment: On estimating elasticities of substition. *Journal of the European Economic Association*, *10*(1), 198–210.

Buffos, L. M., & Chern, W. S. (1984). An energy demand and generalized fuel choice model for the primary metals industry. *The Energy Journal*, *5*(2).

Chen, S. (2017). Interfuel substitution in electricity generation sector: A meta-analysis.

Cho, W. G., Nam, K., & Pagan, J. A. (2004). Economic growth and interfactor/interfuel substitution in korea. *Energy Economics*, *26*(1), 31–50.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. *The review of economics and statistics*, 28–45.

Cohen, A. J., Anderson, H. R., Ostro, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope III, C. A., Romieu, I., Samet, J. M., et al. (2004). Urban air pollution. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, *2*, 1353–1433.

Considine, T. J. (2018). Estimating concave substitution possibilities with non-stationary data using the dynamic linear logit demand model. *Economic Modelling*, *72*, 22–30.

Diewert, W. E. (1973). *Separability and a generalization of the cobb-douglas cost, production and indirect utility functions*. Research Branch, Program Development Service, Department of Manpower and â€¦.

Doucouliagos, C., & Stanley, T. D. (2013). Are all economic facts greatly exaggerated? theory competition and selectivity. *Journal of Economic Surveys*, *27*(2), 316–339.

Du, Z., Lin, B., & Li, M. (2021). Is factor substitution an effective way to save energy and reduce emissions? evidence from china's metallurgical industry. *Journal of Cleaner Production*, *287*, 125531.

Duncan, R. C., & Binswanger, H. P. (1976). Energy sources: Substitutability and biases in australia. *Australian Economic Papers*, *15*(27), 289–301.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315* (7109), 629–634.

Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, *26* (1), 30–55.

Elliott, G., Kudrin, N., & Wüthrich, K. (2022). Detecting p-hacking. *Econometrica*, *90* (2), 887–906.

Furukawa, C. (2019). Publication bias under aggregation frictions: Theory, evidence, and a new correction method. *Evidence, and a New Correction Method (March 29, 2019)*.

Fuss, M. A. (1977). The demand for energy in canadian manufacturing: An example of the estimation of production structures with many inputs. *Journal of econometrics*, *5* (1), 89–116.

Gechert, S., Havranek, T., Irsova, Z., & Kolcunova, D. (2021). Measuring capital-labor substitution: The importance of method choices and publication bias. *Review of Economic Dynamics*.

George, E. I. (2010). Dilution priors: Compensating for model space redundancy. *Borrowing strength: Theory powering applications–a festschrift for lawrence d. brown* (pp. 158–165). Institute of Mathematical Statistics.

Gerber, A., & Malhotra, N. (2008). Do statistical reporting standards affect what is published? publication bias in two leading political science journals. *Quarterly Journal of Political Science*, *3* (3), 313–326.

Griffin, J. M. (1977). Inter-fuel substitution possibilities: A translog application to intercountry data. *International Economic Review*, 755–770.

Hall, V. B. (1983). Industrial sector interfuel substitution following the first major oil shock. *Economics Letters*, *12* (3-4), 377–382.

Hall, V. B. (1986a). Major oecd country industrial sector interfuel substitution estimates, 1960–1979. *Energy Economics*, *8* (2), 74–89.

Hall, V. B. (1986b). Industrial sector fuel price elasticities following the first and second major oil price shocks. *Economics Letters*, *20* (1), 79–82.

Halvorsen, R. (1977). Energy substitution in us manufacturing. *The Review of Economics and Statistics*, 381–388.

Hannah Ritchie, M. R., & Rosado, P. (2020). Co2 and greenhouse gas emissions [https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions]. *Our World in Data*.

Harvey, A. C., & Marshall, P. (1991). Inter-fuel substitution, technical change and the demand for energy in the uk economy. *Applied Economics*, *23*(6), 1077–1086.

He, Y., & Lin, B. (2019). Heterogeneity and asymmetric effects in energy resources allocation of the manufacturing sectors in china. *Energy*, *170*, 1019–1035.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. *Statistical science*, *14*(4), 382–417.

Hossain, A. N., & Serletis, A. (2017). A century of interfuel substitution. *Journal of Commodity Markets*, *8*, 28–42.

Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research.

Iqbal, M. (1986). Substitution of labour, capital and energy in the manufacturing sector of pakistan. *Empirical Economics*, *11*(2), 81–95.

Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.

Jones, C. T. (1995). A dynamic analysis of interfuel substitution in us industrial energy demand. *Journal of Business & Economic Statistics*, 459–465.

Jones, C. T. (1996). A pooled dynamic analysis of interfuel substitution in industrial energy demand by the g-7 countries. *Applied Economics*, *28*(7), 815–821.

Jones, C. T. (2014). The role of biomass in us industrial interfuel substitution. *Energy policy*, *69*, 122–126.

Kamenická, L. (2021). Income inequality and happiness: A meta-analysis.

Kim, B. C., & Labys, W. C. (1988). Application of the translog model of energy substitution to developing countries: The case of korea. *Energy Economics*, *10*(4), 313–323.

Kim, K. (2019). Elasticity of substitution of renewable energy for nuclear power: Evidence from the korean electricity industry. *Nuclear Engineering and Technology*, *51*(6), 1689–1695.

Ko, J., & Dahl, C. (2001). Interfuel substitution in us electricity generation. *Applied Economics*, *33*(14), 1833–1843.

Koetse, M. J., De Groot, H. L., & Florax, R. J. (2008). Capital-energy substitution and shifts in factor demand: A meta-analysis. *Energy Economics*, *30*(5), 2236–2251.

Kranz, S. (2022). Methods matter: P-hacking and publication bias in causal analysis in economics: Comment. *American economic review (forthcoming)*.

Li, J., & Lin, B. (2016). Inter-factor/inter-fuel substitution, carbon intensity, and energy-related co2 reduction: Empirical evidence from china. *Energy Economics*, *56*, 483–494.

Li, J., & Sun, C. (2018). Towards a low carbon economy by removing fossil fuel subsidies? *China Economic Review*, *50*, 17–33.

Lin, B., & Tian, P. (2017). Energy conservation in chinaâ€™s light industry sector: Evidence from inter-factor and inter-fuel substitution. *Journal of Cleaner Production*, *152*, 125–133.

Lin, B., & Xu, M. (2019). Exploring the green total factor productivity of china's metallurgical industry under carbon tax: A perspective on factor substitution. *Journal of Cleaner Production*, *233*, 1322–1333.

Liu, H., & Lin, B. (2017). Energy substitution, efficiency, and the effects of carbon taxation: Evidence from china's building construction industry. *Journal of Cleaner Production*, *141*, 1134–1144.

Liu, K., Bai, H., Yin, S., & Lin, B. (2018). Factor substitution and decomposition of carbon intensity in china's heavy industry. *Energy*, *145*, 582–591.

Ma, C., & Stern, D. I. (2016). Long-run estimates of interfuel and interfactor elasticities. *Resource and Energy Economics*, *46*, 114–130.

Ma, H., Oxley, L., Gibson, J., & Kim, B. (2008). China's energy economy: Technical change, factor demand and interfactor/interfuel substitution. *Energy Economics*, *30*(5), 2167–2183.

Ma, H., Oxley, L., Gibson, J., & Kim, B. (2009). Modeling china's energy consumption behavior and changes in energy intensity. *Environmental Modelling & Software*, *24*(11), 1293–1301.

Magnus, J. R., & Woodland, A. D. (1987). Inter-fuel substitution in dutch manufacturing. *Applied Economics*, *19*(12), 1639–1664.

Miller, J. N. (1993). Tutorial review - outliers in experimental data and their treatment. *Analyst*, *118*(5), 455–461.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, *114*(7), 1518–1523.

Pettersson, F., Söderholm, P., & Lundmark, R. (2012). Fuel switching and climate and energy policies in the european power generation sector: A generalized leontief model. *Energy Economics*, *34*(4), 1064–1073.

Pindyck, R. S. (1979). Interfuel substitution and the industrial demand for energy: An international comparison. *The Review of Economics and Statistics*, 169–179.

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

Serletis, A., Timilsina, G. R., & Vasetsky, O. (2010a). Interfuel substitution in the united states. *Energy Economics*, *32*(3), 737–745.

Serletis, A., Timilsina, G. R., & Vasetsky, O. (2010b). International evidence on sectoral interfuel substitution. *The Energy Journal*, *31*(4).

Shahiduzzaman, M., & Alam, K. (2014). Interfuel substitution in a ustralia: A way forward to achieve environmental sustainability. *Australian Journal of Agricultural and Resource Economics*, *58*(1), 22–42.

Shephard, R. W. (1953). Cost and production functions princeton university press. *Princeton, New Jersey*.

Shin, E. (1981). Inter-energy substitution in korea, 1962-1975.

Söderholm, P. (2001). Fossil fuel flexibility in west european power generation and the impact of system load factors. *Energy economics*, *23*(1), 77–97.

Stanley, T. D. (2005). Beyond publication bias. *Journal of economic surveys*, *19*(3), 309–345.

Stanley, T. D. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and statistics*, *70*(1), 103–127.

Stanley, T. D., & Doucouliagos, H. (2010). Picture this: A simple graph that reveals much ado about research. *Journal of Economic Surveys*, *24*(1), 170–191.

Stanley, T. D., & Doucouliagos, H. (2017). Neither fixed nor random: Weighted least squares meta-regression. *Research synthesis methods*, *8*(1), 19–42.

Stanley, T. D., Jarrell, S. B., & Doucouliagos, H. (2010). Could it be better to discard 90% of the data? a statistical paradox. *The American Statistician*, *64*(1), 70–77.

Steinbuks, J. (2012). Interfuel substitution and energy use in the uk manufacturing sector. *The Energy Journal*, *33*(1).

Steinbuks, J., & Narayanan, B. G. (2015). Fossil fuel producing economies have greater potential for industrial interfuel substitution. *Energy Economics*, *47*, 168–177.

Stern, D. I. (2012). Interfuel substitution: A meta-analysis. *Journal of Economic Surveys*, *26*(2), 307–331.

Suh, D. H. (2016). Interfuel substitution and biomass use in the us industrial sector: A differential approach. *Energy*, ***102***, 24–30.

Taheri, A. A. (1994). Oil shocks and the dynamics of substitution adjustments of industrial fuels in the us. *Applied Economics*, ***26*** (8), 751–756.

Tan, R., & Lin, B. (2020). The influence of carbon tax on the ecological efficiency of china's energy intensive industriesâ€"a inter-fuel and inter-factor substitution perspective. *Journal of environmental management*, ***261***, 110252.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of clinical epidemiology*, ***53*** (2), 207–216.

Urga, G. (1999). An application of dynamic specifications of factor demand equations to interfuel substitution in us industrial energy demand. *Economic Modelling*, ***16*** (4), 503–513.

Urga, G., & Walters, C. (2003). Dynamic translog and linear logit models: A factor demand analysis of interfuel substitution in us industrial energy demand. *Energy Economics*, ***25*** (1), 1–21.

Uri, N. D. (1977). An assessment of interfuel substitution by electric utilities. *Applied Mathematical Modelling*, ***1*** (5), 253–256.

Uri, N. D. (1978). Interfuel substitution possibilities: Short-term prospects. *Applied Energy*, ***4*** (4), 251–260.

Uri, N. D. (1979a). Energy substitution in the uk, 1948–64. *Energy Economics*, ***1*** (4), 241–244.

Uri, N. D. (1979b). Interregional energy substitution in manufacturing in the united states. *Energy Conversion*, ***19*** (1), 25–31.

Uri, N. D. (1981). Forecasting fossil fuel demand. *Electric Power Systems Research*, ***4*** (4), 247–254.

Uri, N. D. (1982a). The electric utility demand for energy in the united states. *empirical economics*, ***7*** (1), 75–92.

Uri, N. D. (1982b). Energy demand and interfuel substitution in the united kingdom. *Socio-Economic Planning Sciences*, ***16*** (4), 157–162.

Uri, N. D. (1982c). The industrial demand for energy. *Resources and Energy*, ***4*** (1), 27–57.

van Aert, R. C., & van Assen, M. (2021). Correcting for publication bias in a meta-analysis with the p-uniform* method. *Working Paper*.

Wang, A., & Lin, B. (2020). Structural optimization and carbon taxation in china's commercial sector. *Energy Policy*, ***140***, 111442.

Wang, X., Bai, M., & Xie, C. (2019). Investigating co2 mitigation potentials and the impact of oil price distortion in china's transport sector. *Energy Policy*, *130*, 320–327.

Wang, X., & Lin, B. (2017). Factor and fuel substitution in china's iron & steel industry: Evidence and policy implications. *Journal of cleaner production*, *141*, 751–759.

Welsch, H., & Ochsen, C. (2005). The determinants of aggregate energy use in west germany: Factor substitution, technological change, and trade. *Energy Economics*, *27*(1), 93–111.

Yang, M., Fan, Y., Yang, F., & Hu, H. (2014). Regional disparities in carbon dioxide reduction from china's uniform carbon tax: A perspective on interfactor/interfuel substitution. *Energy*, *74*, 131–139.

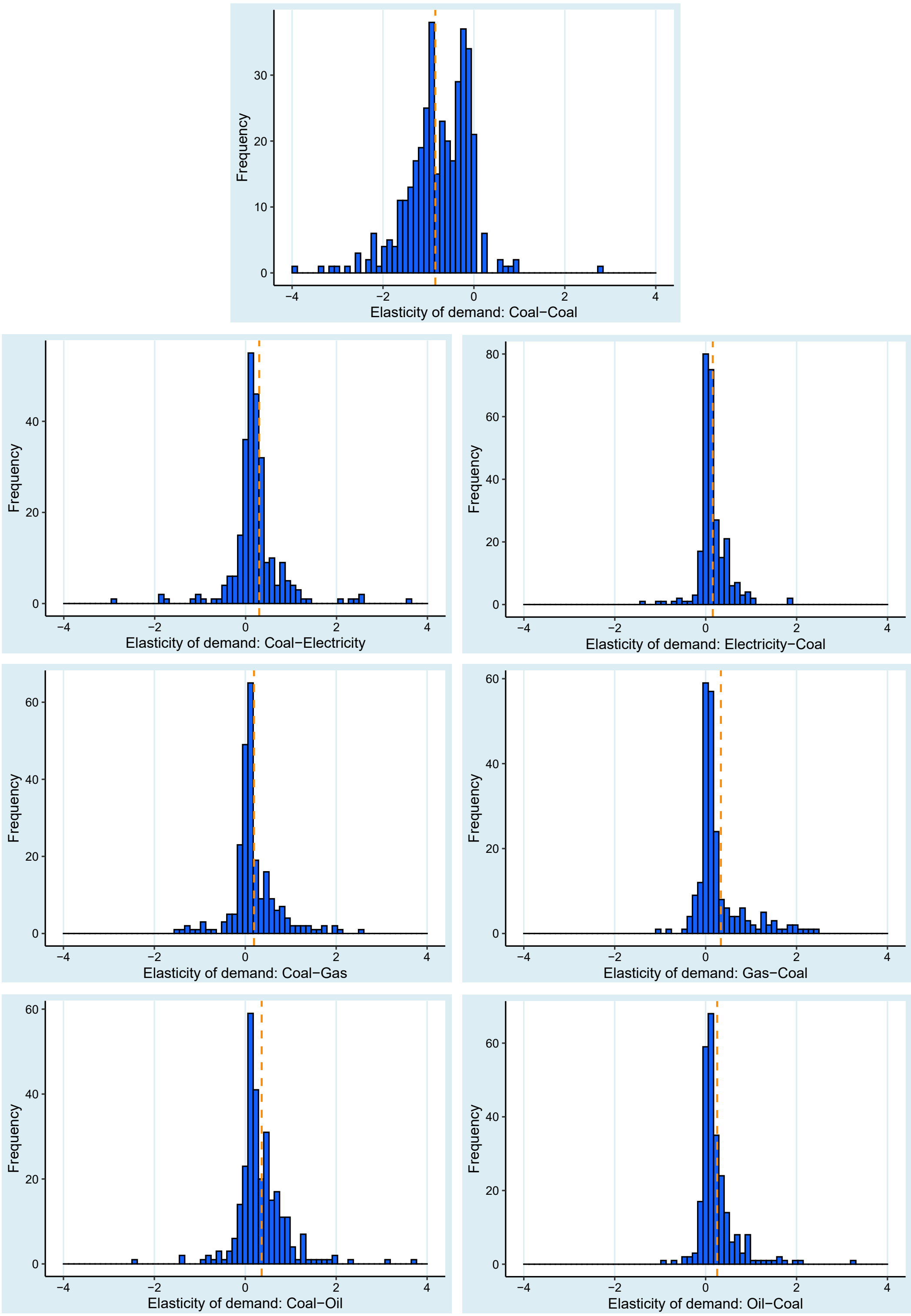
Zeugner, S., & Feldkircher, M. (2015). Bayesian model averaging employing fixed and flexible priors: The bms package for r. *Journal of Statistical Software*, *68*, 1–37.

Zhang, Z., & Lin, B. (2019). Energy conservation and emission reduction of chinese cement industry: From a perspective of factor substitutions. *Emerging Markets Finance and Trade*, *55*(5), 967–979.

Appendix A

Data

Figure A.1: Distribution of all the reported estimates



Notes: The figure shows the distribution of the reported estimates, including estimates for which it was impossible to calculate standard errors. The vertical line denotes the mean elasticity of demand for each subset. Estimates smaller than -4 and greater than 4 are excluded from the figure for the sake of readability.

Table A.1: List of all reviewed studies

Andrikopoulos et al. (1989)	Halvorsen (1977)
Ma et al. (2009)	Atkinson and Halvorsen (1976)
Harvey and Marshall (1991)	Ma and Stern (2016)
Bello et al. (2020)	He and Lin (2019)
Magnus and Woodland (1987)	Borges and Pereira (1992)
Iqbal (1986)	Pettersson et al. (2012)
Cho et al. (2004)	Jones (1995)
Pindyck (1979)	Considine (2018)
Kim and Labys (1988)	Shahiduzzaman and Alam (2014)
Duncan and Binswanger (1976)	Kim (2019)
Shin (1981)	Fuss (1977)
Ko and Dahl (2001)	Söderholm (2001)
Griffin (1977)	Li and Lin (2016)
Steinbuks (2012)	Hall (1983)
Lin and Tian (2017)	Steinbuks and Narayanan (2015)
Hall (1986b)	Ma et al. (2008)
Suh (2016)	Hall (1986a)
Wang and Lin (2020)	Taheri (1994)
Urga (1999)	Urga and Walters (2003)
Wang et al. (2019)	Uri (1978)
Uri (1979a)	Wang and Lin (2017)
Yang et al. (2014)	Jones (1995)
Jones (1996)	Bu□os and Chern (1984)
Liu and Lin (2017)	Lin and Xu (2019)
Uri (1981)	Serletis et al. (2010b)
Du et al. (2021)	Jones (2014)
Hossain and Serletis (2017)	Uri (1977)
Zhang and Lin (2019)	Uri (1979a)
Uri (1982b)	Liu et al. (2018)
Uri (1978)	Serletis et al. (2010a)
Uri (1979b)	Uri (1982a)
Uri (1982c)	Tan and Lin (2020)
Li and Sun (2018)	

Notes: The table shows list of all reviewed studies which reported empirical estimates of the elasticity of demand, including those which did not report estimates' standard errors, statistics sufficient to compute standard errors or standard errors of other parameters, based on which the standard errors could be derived.

Table A.2: The mean and median reported elasticity (including all the estimates)

Subset	Observations	Unweighted mean	Weighted mean	Median
Coal-Coal	379	-0.85	-0.63	-0.72
Coal-Electricity	268	0.30	0.19	0.18
Coal-Gas	247	0.19	0.12	0.10
Coal-Oil	285	0.36	0.27	0.24
Electricity-Coal	270	0.15	0.09	0.08
Gas-Coal	255	0.34	0.27	0.10
Oil-Coal	263	0.25	0.24	0.12

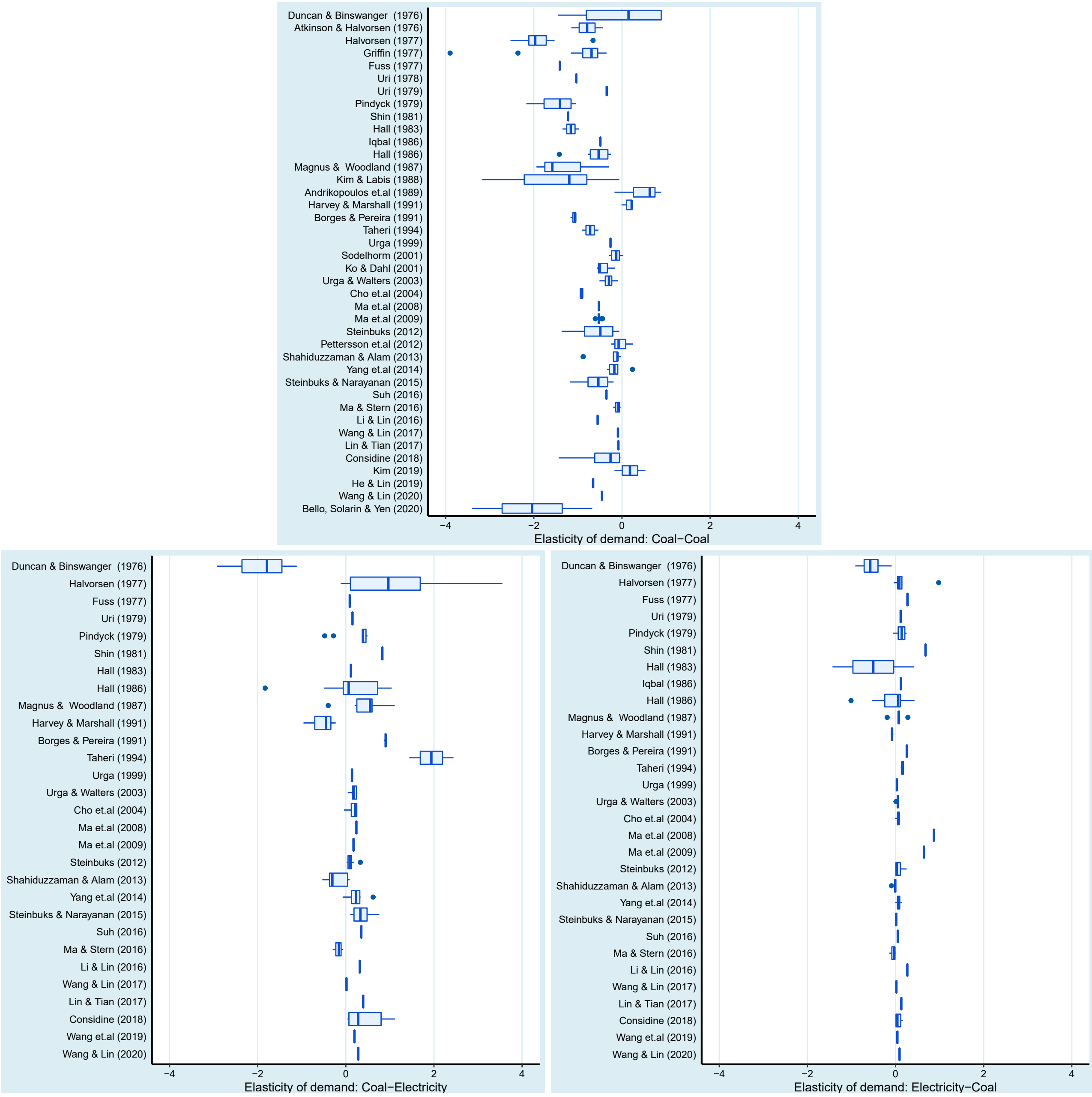
Notes: The table reports the mean and median reported estimate of the elasticity of substitution for each subset, including estimates for which it was impossible to calculate standard errors. Weighted = estimates are weighted by the inverse of the number of observations reported per study.

Table A.3: Summary statistics of the reported elasticity (including all the estimates)

Subset	Observations	St.Dev.	Min	Max
Coal-Coal	379	0.97	-9.69	2.78
Coal-Electricity	268	0.99	-5.29	6.59
Coal-Gas	246	0.53	-1.55	2.52
Coal-Oil	285	0.64	-2.40	5.52
Electricity-Coal	270	0.32	-1.43	1.91
Gas-Coal	255	0.78	-1.06	5.99
Oil-Coal	263	0.49	-0.92	4.36

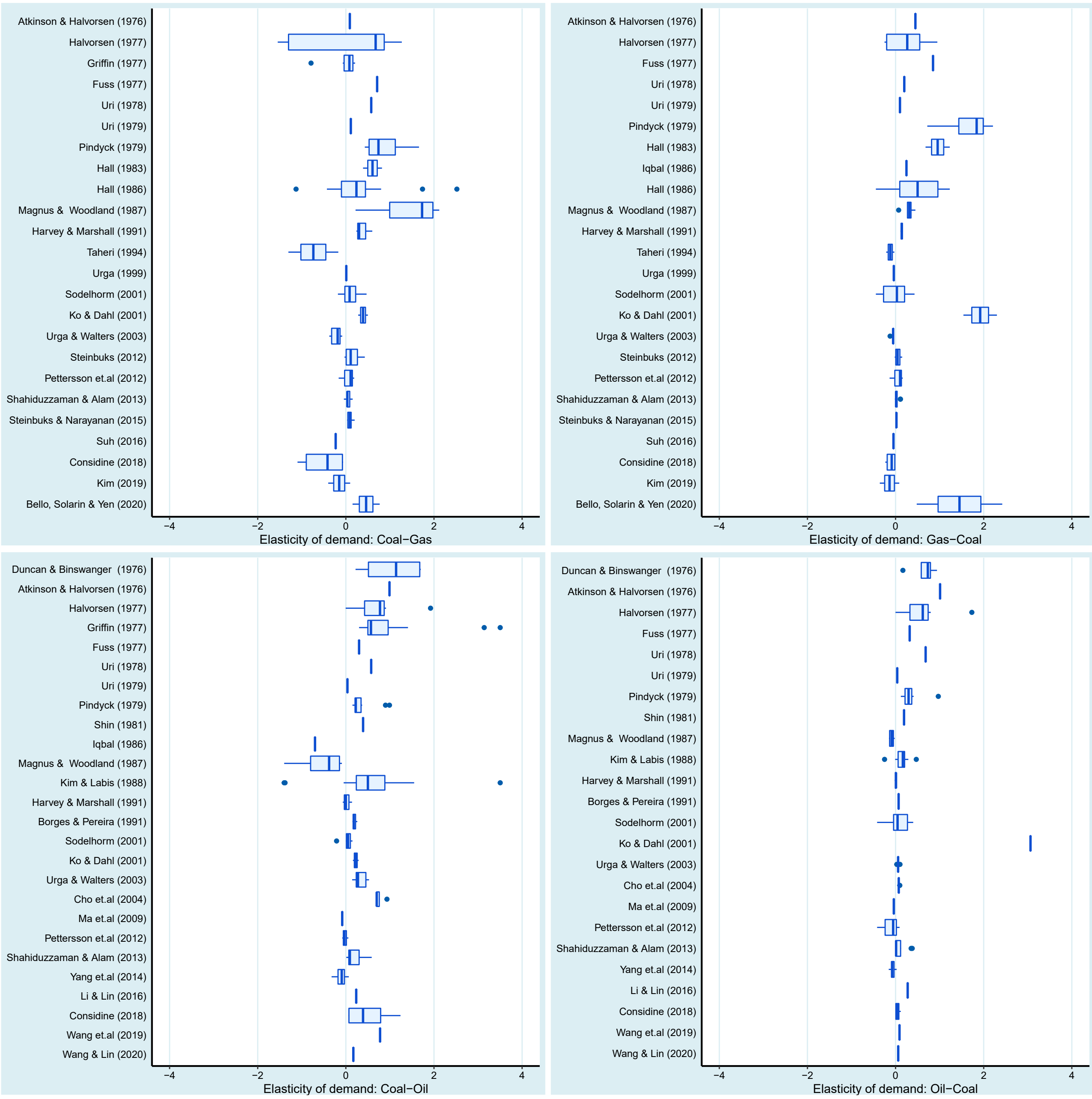
Notes: The table reports summary statistics of the reported elasticity for each subset, including estimates for which it was impossible to calculate standard errors. St.Dev. = standard deviation.

Figure A.2: Boxplots show that estimates vary both within and across studies



Notes: The box length represents the interquartile range. The vertical line inside the box denotes the median value. The whiskers denote the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. Studies are sorted by year of publication from the oldest to the most recent. Estimates smaller than -4 and greater than 4 are excluded from the figure for the ease of exposition but included in all statistical tests.

Figure A.3: Boxplots show that estimates vary both within and across studies

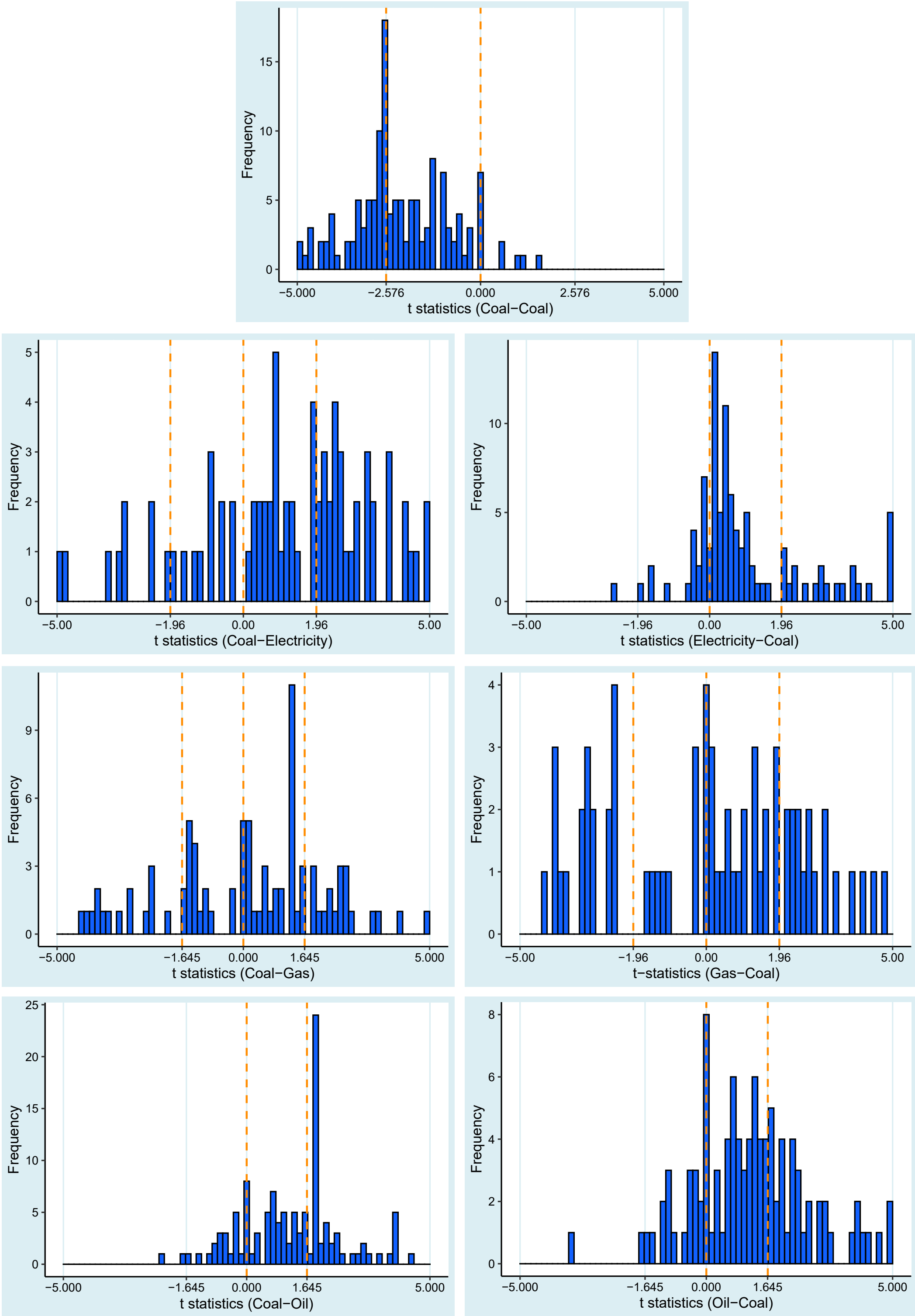


Notes: The box length represents the interquartile range. The vertical line inside the box denotes the median value. The whiskers denote the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. Studies are sorted by year of publication from the oldest to the most recent. Estimates smaller than -4 and greater than 4 are excluded from the figure for the ease of exposition but included in all statistical tests.

Appendix B

Publication bias

Figure B.1: Distribution of the t-statistics of the reported estimates



Notes: The figure shows the distribution of the t-statistics of the reported estimates. Vertical lines denote the cut-offs that were passed to the Selection model, developed by Andrews and Kasy (2019) to detect publication bias. Values of t-statistic equal to 1.645, 1.96 and 2.576 translate to the commonly used p-values of 0.1, 0.05 and 0.001, respectively. The results of the model are reported in Table 4.2. T-statistics smaller than -5 and greater than 5 are excluded from the figure for the sake of readability.

Table B.1: Specification test for the Selection model developed by Andrews and Kasy (2019).

Subset	Trans.	Corr.
Coal-Coal	level	0.13
		[0.067,0.617]
	log	0.48
		[-0.665, 0.840,]
Coal-Electricity	level	0.31
		[-0.374, 0.729]
	log	0.58
		[0.445, 0.783]
Coal-Gas	level	0.07
		[-0.078, 0.308]
	log	0.62
		[0.470, 0.718]
Coal-Oil	level	-0.06
		[-0.170, 0.388]
	log	0.44
		[0.353, 0.705]
Electricity-Coal	level	-0.49
		[-0.680, -0.120]
	log	0.69
		[0.558, 0.781]
Gas-Coal	level	-0.03
		[-0.178, 0.223]
	log	0.842
		[0.757, 0.902]
Oil-Coal	level	-0.29
		[-0.434, 0.147]
	log	0.64
		[0.521, 0.850]

Notes: The table shows, following Kranz (2022), the correlation coe□cient between the logarithm of the absolute value of the estimated inverse elasticity and the logarithm of the corresponding standard error, weighted by the inverse publication probability estimated by the Selection model (Andrews and Kasy, 2019). If the assumptions of the model hold, the correlation in both levels and logarithmic transformation should be zero. Bootstrap confidence intervals are reported in the square brackets.

Table B.2: Caliper tests

Coal-Coal subset			
Threshold for t-statistic: -2.576	width: 0.4	width: 0.5	width: 0.6
Share of estimates below caliper	0.1374$^{\square\square\square}$	0.1754$^{\square\square\square}$	0.1986$^{\square\square\square}$
	(0.0194)	(0.0239)	(0.0240)
Observations	19	31	39
Threshold for t-statistic: 0	width: 1.3	width: 1.4	width: 1.5
Share of estimates below caliper	-0.4239$^{\square\square\square}$	-0.3873$^{\square\square\square}$	-0.3618$^{\square\square\square}$
	(0.1152)	(0.1004)	(0.0922)
Observations	37	40	42
Coal-Electricity subset			
Threshold for t-statistic: 1.96	width: 0.1	width: 0.2	width: 0.3
Share of estimates above caliper	-0.2916	-0.2743	-0.1950$^{\square\square\square}$
	(0.1243)	(0.1002)	(0.0747)
Observations	5	7	10
Threshold for t-statistic: 0	width: 1	width: 1.1	width: 1.2
Share of estimates above caliper	0.3991$^{\square\square\square}$	0.4075$^{\square\square\square}$	0.3999$^{\square\square\square}$
	(0.1765)	(0.1640)	(0.1405)
Observations	21	22	24
Coal-Gas subset			
Threshold for t-statistic: 1.65	width: 0.1	width: 0.2	width: 0.3
Share of estimates above caliper	-0.2899	-0.2899	-0.22008$^{\square}$
	(0.2034)	(0.2034)	(0.0907)
Observations	3	3	10
Threshold for t-statistic: 0	width: 1.1	width: 1.2	width: 1.3
Share of estimates above caliper	0.3915$^{\square\square\square}$	0.2684$^{\square\square\square}$	0.1236$^{\square\square\square}$
	(0.2011)	(0.1930)	(0.1507)
Observations	25	26	30
Coal-Oil subset			
Threshold for t-statistic: 1.65	width: 0.1	width: 0.2	width: 0.3
Share of estimates above caliper	-0.3706	-0.2801	-0.0688$^{\square\square\square}$
	(0.1214)	(0.0961)	(0.0362)
Observations	5	9	34
Threshold for t-statistic: 0	width: 1	width: 1.1	width: 1.2
Share of estimates above caliper	0.4646$^{\square\square\square}$	0.4147$^{\square\square\square}$	0.4101$^{\square\square\square}$
	(0.1389)	(0.1163)	(0.1070)
Observations	44	49	51

Notes: The table reports results for caliper tests, developed by Gerber and Malhotra (2008). The tests compare the relative frequency of estimates above and below a certain threshold for the t-statistic. Standard errors are shown in parentheses and clustered at the study-level. $^{\square}$, $^{\square\square}$ and $^{\square\square\square}$ denote statistical significance at 1%, 5% and 10%.

Table B.3: Caliper tests

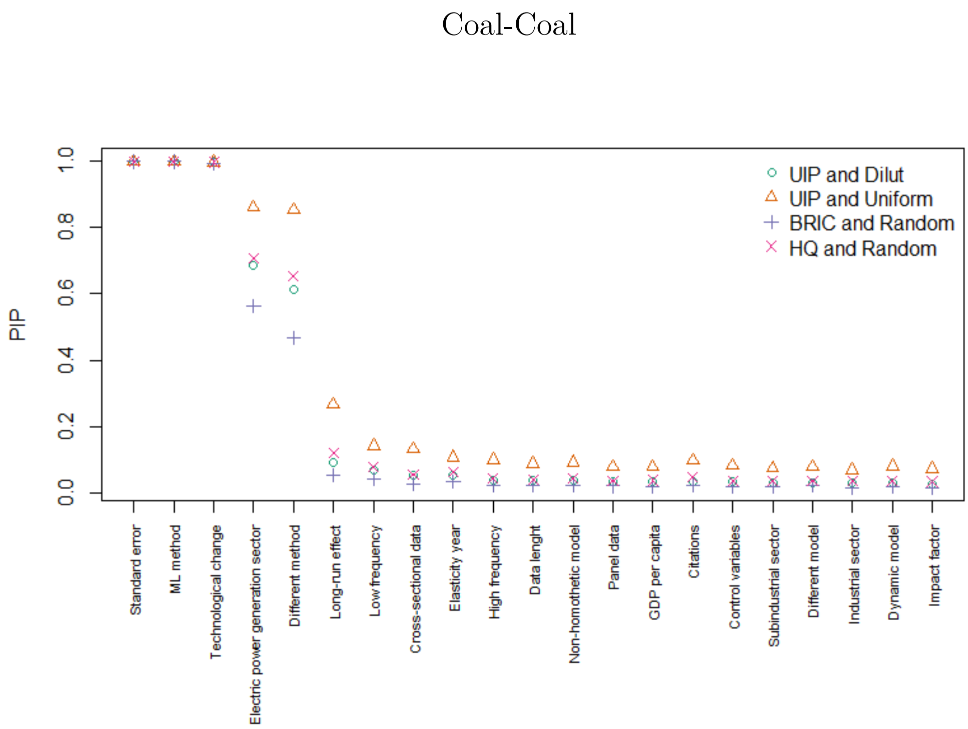
Electricity-Coal subset			
Threshold for t-statistic: 1.96	width: 0.1	width: 0.2	width: 0.3
Share of estimates above caliper	-0.1340**	-0.0981***	-0.0879***
	(0.0900)	(0.0628)	(0.0514)
Observations	7	10	12
Threshold for t-statistic: 0	width: 1	width: 1.1	width: 1.2
Share of estimates above caliper	0.3910***	0.4037***	0.3893***
	(0.2148)	(0.1964)	(0.1396)
Observations	18	19	23
Gas-Coal subset			
Threshold for t-statistic: 1.96	width: 0.2	width: 0.3	width: 0.4
Share of estimates above caliper	-0.3571	-0.3084	-0.2816*
	(0.1237)	(0.1019)	(0.0857)
Observations	4	6	8
Threshold for t-statistic: 0	width: 1	width: 1.1	width: 1.2
Share of estimates above caliper	0.3348**	0.3250***	0.1149***
	(0.2157)	(0.1893)	(0.1449)
Observations	20	21	25
Oil-Coal subset			
Threshold for t-statistic: 1.65	width: 0.2	width: 0.3	width: 0.4
Share of estimates above caliper	-0.1874	-0.1791**	-0.1864*
	(0.1312)	(0.0862)	(0.0698)
Observations	6	12	17
Threshold for t-statistic: 0	width: 1	width: 1.1	width: 1.2
Share of estimates above caliper	0.3516***	0.3082***	0.3082***
	(0.1348)	(0.1169)	(0.1169)
Observations	45	49	49

Notes: The table reports results for caliper tests developed by Gerber and Malhotra (2008). The tests compare the relative frequency of estimates above and below a certain threshold for the t-statistic. Standard errors are shown in parentheses and clustered at the study-level. *, ** and *** denote statistical significance at 1%, 5% and 10%.

Appendix C

Heterogeneity

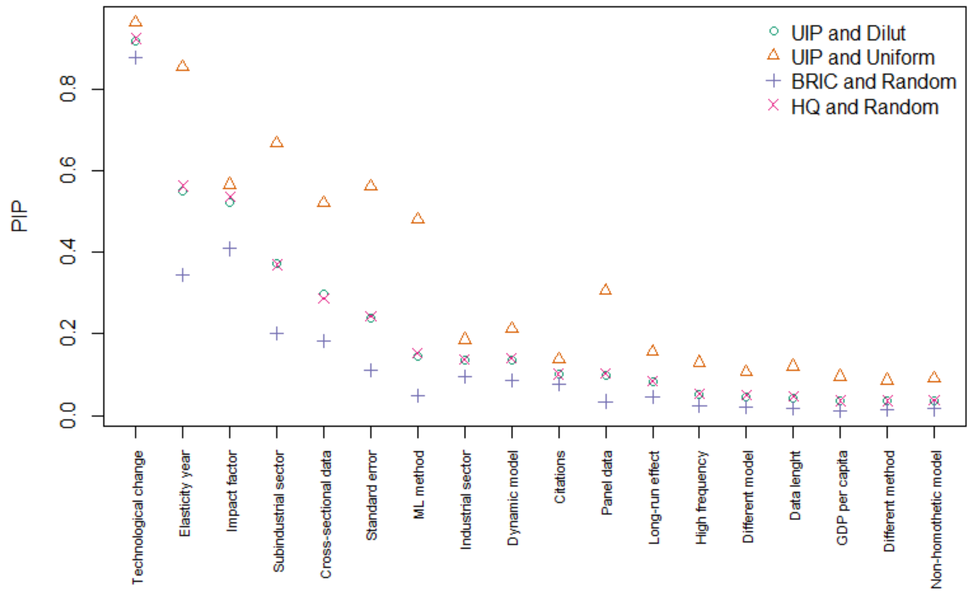
Figure C.1: PIP across different prior settings, own-price elasticity for coal



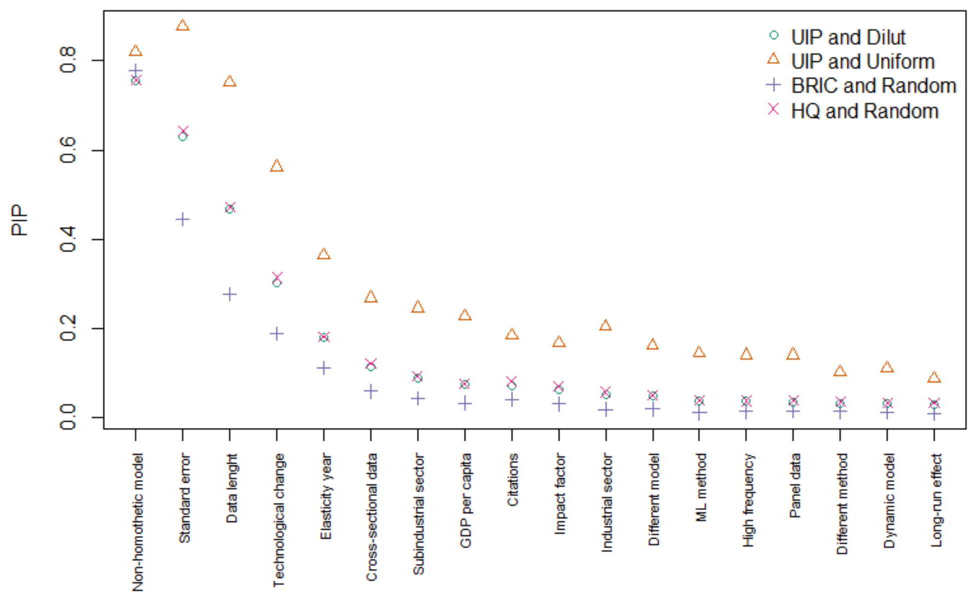
Notes: The figure shows PIP across different prior settings. The response variable is the estimate of the own-price elasticity of demand for coal. UIP and Dilut = prior due to George (2010); UIP and Uniform = prior due to Eicher et al. (2011); BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for the model space, meaning that each model size has equal prior probability; HQ prior asymptotically simulates the Hannan-Quinn criterion. UIP = unit information prior; PIP = posterior inclusion probability.

Figure C.2: PIP across different prior settings, coal-electricity substitution

(a) Coal-Electricity



(b) Electricity-Coal



Notes: The figure shows PIP across different prior settings. The response variable is the estimate of the cross-price elasticity of demand between coal and electricity at the top and the estimate of the cross-price elasticity of demand between electricity and coal at the bottom. UIP and Dilut = prior due to George (2010); UIP and Uniform = prior due to Eicher et al. (2011); BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for the model space, meaning that each model size has equal prior probability; HQ prior asymptotically simulates the Hannan-Quinn criterion. UIP = unit information prior; PIP = posterior inclusion probability.

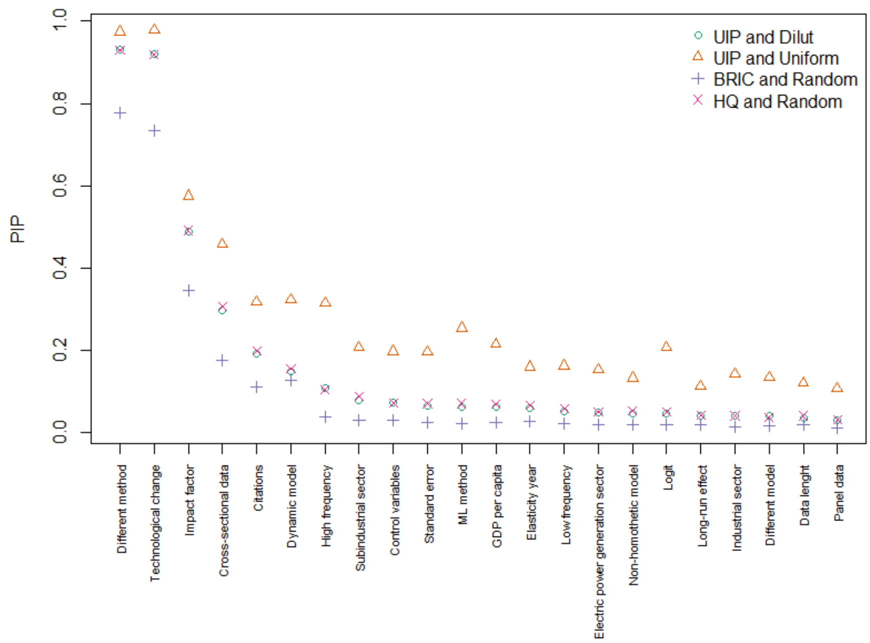
Table C.1: Linear tests suggest that publication bias is driven by the industry-level estimates of the cross-price elasticity of demand between gas and coal

Panel A	OLS	BE
Standard error	4.8748**	3.5067***
Publication bias	(1.4157)	(0.5539)
Constant	-0.1554	-0.0560
Mean beyond bias	(0.0914)	(0.0588)
Observations	49	49
Panel B		
Standard error	-0.0408***	0.7190
Publication bias	(0.0070)	(0.4010)
Constant	0.5385***	0.4152**
Mean beyond bias	(0.1320)	(0.1311)
Observations	55	55

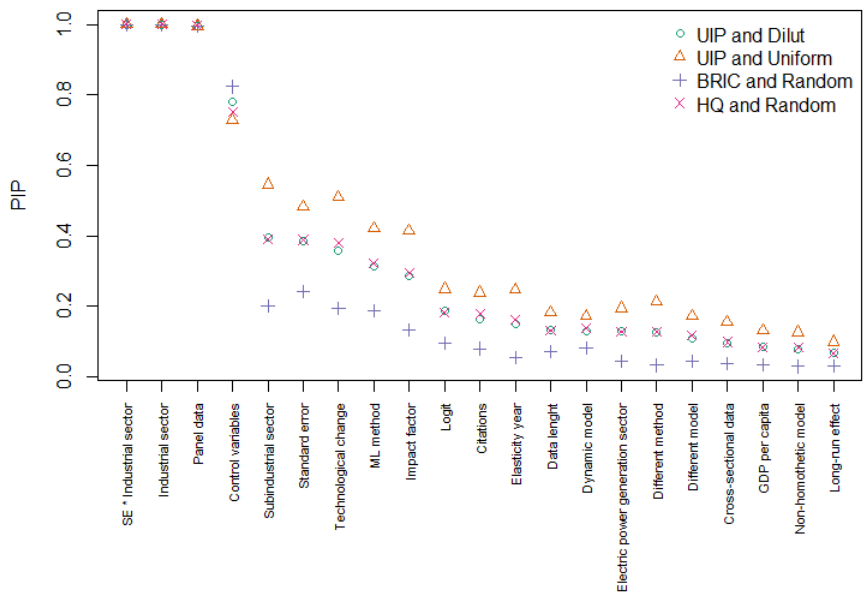
Notes: The table reports results of the regression $\eta_{ij} = \eta_0 + \gamma SE(\eta_{ij}) + \ _{ij}$. η_{ij} and SE(η_{ij}) stand for the i^{th} estimate of the cross-price elasticity of demand between gas and coal and its standard error from the j^{th} study. OLS = ordinary least squares; BE = between-e□ects model. The standard errors are clustered at the study level and reported in parentheses. *, ** and *** denote statistical significance at 1%, 5% and 10% level. Panel A contains only the industry-level estimates. Panel B excludes the industry-level estimates.

Figure C.3: PIP across different prior settings, coal-gas substitution

(a) Coal-Gas



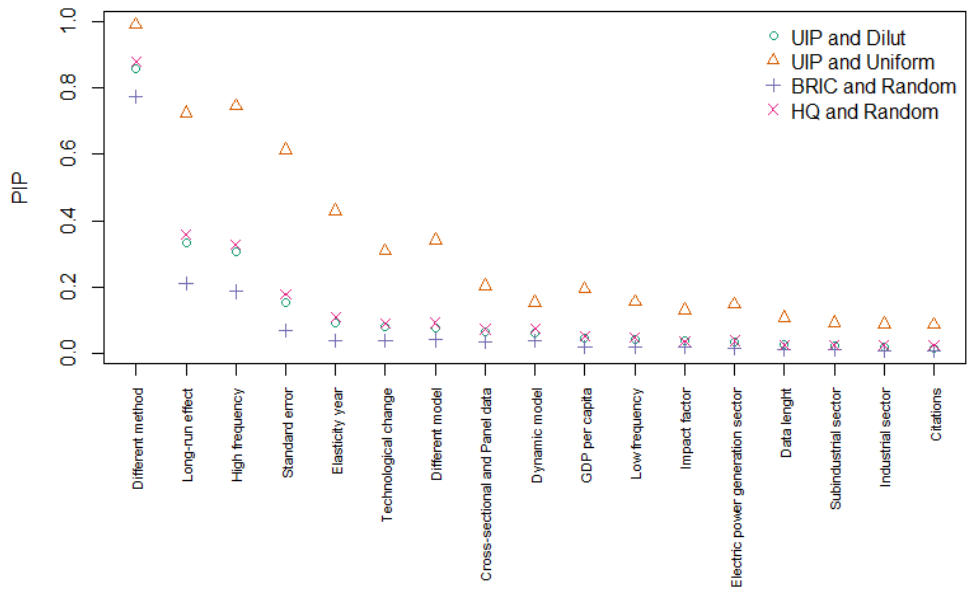
(b) Gas-Coal



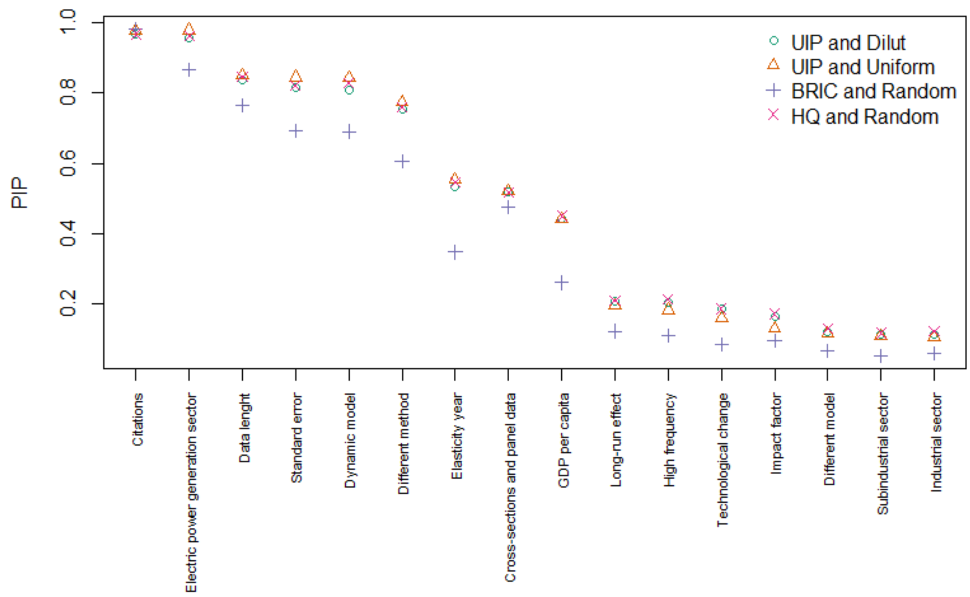
Notes: The figure shows PIP across different prior settings. The response variable is the estimate of the cross-price elasticity of demand between coal and gas at the top and the estimate of the cross-price elasticity of demand between gas and coal at the bottom. UIP and Dilut = prior due to George (2010); UIP and Uniform = prior due to Eicher et al. (2011); BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for the model space, meaning that each model size has equal prior probability; HQ prior asymptotically simulates the Hannan-Quinn criterion. UIP = unit information prior; PIP = posterior inclusion probability.

Figure C.4: PIP across different prior settings, coal-oil substitution

(a) Coal-Oil



(b) Oil-Coal



Notes: The figure shows PIP across di erent prior settings. The response variable is the estimate of the cross-price elasticity of demand between coal and oil at the top and the estimate of the cross-price elasticity of demand between oil and coal at the bottom. UIP and Dilut = prior due to George (2010); UIP and Uniform = prior due to Eicher et al. (2011); BRIC and Random = the benchmark g-prior for parameters with the beta-binomial model prior for the model space, meaning that each model size has equal prior probability; HQ prior asymptotically simulates the Hannan-Quinn criterion. UIP = unit information prior; PIP = posterior inclusion probability.