Received: 4 October 2018 | Revised: 27 February 2019 | Accepted: 8 March 2019
DOI: 10.1111/exsy.12423

ORIGINAL ARTICLE

WILEY Expert Systems

Bargains in the diamond market? How to take advantage from online information

Angel Arcos-Vargas[1] | Fernando Núñez[1] | Antonio J. Tallón-Ballesteros[2]

[1] Department of Industrial Organization and Business Management I, University of Seville, Seville, Spain
[2] Department of Languages and Computer Systems, University of Seville, Seville, Spain

Correspondence
Antonio J. Tallón-Ballesteros, Department of Languages and Computer Systems, University of Seville, Reina Mercedes Avenue. 41012 Seville, Spain.
Email: atallon@us.es

Abstract

This paper empirically analyses the price of diamonds from different perspectives. The sample data contains detailed information on more than 165,000 diamonds certified by Gemological Institute of America, which can be purchased since July 2016 from a large online diamond supplier. Our empirical analysis allows to extend the classic 4Cs model in several directions as follows: (a) the classical 4Cs model is insufficient in order to explain the price, so that other attributes, such as polishing, symmetry, and fluorescence, must be incorporated to accurately estimate the price of a particular piece; (b) the large sample size allows us to analyse specific price ranges characterized by high market activity; (3) a very detailed analysis has been conducted for round diamonds, incorporating interactions between attributes within the regression model and studying the marginal rates of substitution between diamond weight and alternative characteristics given a certain price. Our model is able to detect diamonds that could constitute a clear market opportunity for buyers. This model can help to select efficient demand options taking advantage of the information available on the website, improving the overall efficiency of exchange in this sector.

KEYWORDS

analytics, big data, empirical, market characteristics, multiple linear regression, regression

1 | INTRODUCTION

For more than 3,000 years, diamonds have been surrounded by a halo of mystery, superstition, and intrigue; extraordinary properties have been attributed to them. In fact, they have been considered a symbol of power, strength, courage, command, intelligence, and, even, manhood. Diamonds also served as curative elements. People were convinced that either their proximity to the diseased organ or their ingestion in the form of diamond dust healed; the ingestion of diamond dust led to the death of King Frederick II of Prussia in 1250. Truly, the word "diamond" comes from the Greek term *adamanten*, which means invincible.

This merchandise has not gone unnoticed by economists, amongst other reasons, because it is a commodity that breaks certain market rules. The first reference we find on the behaviour of diamonds corresponds to Pier de Jean Olivi (1248–1298), who in his *Tractatus de Contractibus*, enunciates and solves the paradox of water and diamonds, which five centuries later was popularized by Adam Smith, 1786 (Ekelund & Thornton, 2011). Currently, diamond supply is on the rise due to the increase in the price of this commodity. This growing trend is shown, for example, in the article entitled *Historical Diamond Prices Per Carat from 1960 to 2016 in U.S. Dollars*,[1] which has been published by the statistics web portal "*Statista*." Moreover, the analyst Paul Zimnisky has predicted an increase in the price for 2018.[2] Indeed, the evolution of the price is fundamental

[1]https://www.statista.com/statistics/279053/worldwide-sales-of-polished-diamonds.

[2]http://www.cbc.ca/news/canada/north/nwt-diamond-prices-2018-1.4477796.

in this business, and proof of this is that a new patent for determining the market value of a diamond (with number US-20170140445) has been issued by the World Intellectual Property Organization in the year 2017.

As we will see, the existing literature on the diamond market can be classified into three blocks: research works that analyse, in a general way, the economic structure and functioning of the diamond market, studies that use different databases in order to explain the diamond price determinants, and works that compare the diamond price evolution with the behaviour of other commodities such as gold, shares, or bonds; our article fits into these last two groups.

Most of the market structures (monopoly, oligopoly, and perfect competition) are present in the diamond business conditioning price determination. Furthermore, colonial policy of the states and anticompetitive practices have characterized the situation and evolution of this commodity. These particular characteristics have led many authors to study in depth this merchandise, authors such as Von Saldern (1992), Spar (2006), Bergenstock, Deily, and Taylor (2006), Gupta, Polonsky, Woodside, and Webster (2010), Levin and Sheveleva (2010), Lee, Caudill, and Mixon (2014), and Grynberg and Mbayi (2015).

One important aspect analysed in literature on this topic refers to the determination of the diamond price based on its main characteristics, namely “carat weight”: diamonds are measured by carats; one carat is equal to 0.2 gr; “shape”: amongst the main shapes, we find round, princess, oval, marquise, emerald, asscher, and heart; “colour”: to measure the colour of a diamond, the industry has adopted the diamond scale of the Gemological Institute of America (GIA), with grades ranging from D (*highest grade*) to Z (*lowest grade*). The GIA scale defines the five following groups: colourless (D–F) diamonds, almost colourless (G–J), slightly light yellow (K–M), very light yellow (N–R), and light yellow (S–Z); “Cut”: diamond's cut grade is about how well a diamond's facets interact with light. GIA classifies diamonds with a cut grade ranging from excellent to poor by taking into account different parameters, such as the diamond's proportions, culet size, girdle thickness, polish, and symmetry descriptions; and, finally, “clarity”: which refers to the tiny natural imperfections that are present in all of them. The less inclusions, the better the clarity, and they are classified as flawless (FL, IF), double very slightly included (VVS1, VVS2), very slightly included (VS1, VS2), slightly included (SI1, SI2), and included (I1, I2, I3).

These four characteristics for each given shape—carat weight, colour, clarity, and cut grade—normally known as the 4Cs model, are the most important in order to explain the price of a diamond; however, there are other features that are also taken into account when the price is determined, which are as follows: “polishing”: refers to the degree of smoothness of the faces of a diamond. GIA grades range from excellent (no visible defects under a glass-10) to poor (defects visible to the naked eye); “symmetry”: indicates how the faces of a diamond line up. The degrees of GIA range from excellent (defects of symmetry not visible under a glass-10) to poor (defects visible to the naked eye); “fluorescence”: it is a feature that causes some diamonds to appear to change colour when exposed to ultraviolet light. The degrees of GIA fluorescence vary from very strong to none; and “certification”: diamond certification is a qualification report describing the characteristics and quality of the gemstone. In each certificate, an identification number is given to the corresponding stone.

The business of diamonds has traditionally presented two main drawbacks: the difficulty of knowing and assessing the physical characteristics of diamonds and the fact that market prices and supply and demand functions have been practically unknown, because exchanges have been negotiated individually (often through batches of diamonds where elements of different quality and price were mixed). The first initiative to provide more transparency to the diamond market is known as the Rapaport report,[3] which began to be published in 1978. Since then, this report has provided guidance prices for each possible combination of diamond attributes, which is used as a starting point for negotiation. The dissemination of this report was limited to professional jewellers, not being accessible to the final consumers. However, the entry of operators in the online market (since 2001) has completely changed the degree of information that buyers can have; at present, prices and features of hundreds of thousands of diamonds are available in real time for buyers and suppliers.

The more abundant information on the diamond business has also allowed new academic research. Different online sources have been mentioned in the literature. For instance, Scott and Yelowitz (2010) study three data sources: amazon.com, bluenile.com, and uniondiamond.com, Renneboog and Spaenjers (2012) use data from sothebys.com and christies.com, Auer (2014) and Auer and Schuhmacher (2013) analyse data from polishedprices.com, and Wolff (2015, 2016) employ data from info-diamond.com. Likewise, there are three modalities to access the data of the different online platforms: (a) Privileged: where the data are not for general public use and an agreement with international auction companies is needed—this is the case of sothebys.com and christies.com; (b) To be paid: an annual subscription is required to be able to access the data of the web page—polishedprices.com; and (c) Freely accessible: free disposal of data—amazon.com, bluenile.com, uniondiamond.com, and info-diamond.com. For our research, we have only explored freely accessible data, and within the available ones, we have selected Blue Nile because it is, in our opinion, the widest and best structured web page.

Large databases allow analysing the determinants of the diamond price. For instance, Scott and Yelowitz (2010) analyse a sample of more than 140,000 diamonds obtained in 2005 from three online retailers: Blue Nile, Union, and Amazon. Their empirical model is based on a multiple regression that uses the attributes carat weight, colour, clarity, and cut grade as explanatory variables; they do not consider the rest of the attributes previously mentioned (polishing, symmetry, fluorescence, or certificate). Besides obtaining significant coefficients in the regression, they found a certain anomaly in the prices when the weight approached certain values. One possible explanation for this price abnormality is that there is

[3]See Rapaport Diamond Price Statistics, Annual Report, 2016.

TABLE 1 Review of the literature

Authors (year)	Historical evolution	Market structure	Pricing	Investment vehicle	Price discontinuity	Regression model	Methodology
Von Saldern (1992)		✓	✓				Nonlinear optimization
Spar (2006)	✓	✓					Market description
Bergenstock et al. (2006)	✓	✓					Cointegration analisys
Gupta et al. (2010)	✓	✓					Compare two study cases
Scott and Yelowitz (2010)			✓		✓	✓	Logarithmic regression
Levin and Sheveleva (2010)	✓	✓					Market description
Renneboog and Spaenjers (2012)				✓		✓	Hedonic regression
Auer (2014)				✓			Multivariate Garch model
Auer and Schuhmacher (2013)				✓			Multivariate Garch model
Lee et al. (2014)	✓					✓	Hedonic regression
Wolff (2015)			✓			✓	Hedonic regression
Wolff (2016)			✓			✓	DEA

Abbreviation: DEA, data envelopment analysis.

an integer number effect; for instance, consumers perceive that there is a categorical difference between diamonds smaller than one carat and those of one carat or a little more.

For their part, Lee et al. (2014) amplify the previous investigation by analysing the effect on price of the certifying institute in a sample of 169,483 observations extracted from the website PriceScope.com in February 2013; the certifying institutes are the American Gemological Society, the GIA, and the European Gemological Laboratory. The model proposed by these authors expands the previous one of Scott and Yelowitz by adding cut grade (from ideal to poor), polishing (from ideal to poor), fluorescence (from very strong to none), and certification attributes. They observe that diamonds classified by European Gemological Laboratory present a price discount; that is, the certifying office produces an additional risk factor that pushes the price down.

In the line of price analysis, Wolff (2015) investigates the causes of price dispersion in apparently homogeneous diamonds. He analyses the online diamond market using a sample of 159,253 certified diamonds offered in November 2012 on the website at www.diamonts-infos.com. The author includes, in a multiple regression model on the diamond price, the 4Cs explanatory variables along with the polishing, symmetry, fluorescence, and certification attributes. His model explains 97% of the price variation, concluding further that both, prices and price dispersion, increase significantly with quality. An explanation of the positive relation between quality and price dispersion might be that buyers of high-quality diamonds are willing to pay a higher price and even more so when they perceive high prices as a signal of power and exclusivity. In another study, Wolff (2016) studies the relative efficiency of buyers and sellers in the diamond market by using a data envelopment analysis model with double-frontier adjustment. He identifies the groups of efficient sellers and buyers on the price–weight space. For each diamond, the distance to the frontier allows to determine how inefficient the purchase (or sale) of that diamond is. The main finding of this research is that, in general, the bargaining power of buyers is lower than that of the seller.

Another interesting set of articles in this research field, although far from the scope of our study, is the one that considers diamonds as a potential investment asset, as an alternative to the traditional commodities gold and silver. This topic is developed, for instance, in Baur and Lucey (2010), Auer (2014), Auer and Schuhmacher (2013), Joy (2011), and Renneboog and Spaenjers (2012), who study the market of diamonds and gems as an investment vehicle in the early years of this century (1999–2010), where two international crises occurred.[4] They conclude that although diamond returns have been below those of gold, diamonds have outperformed the stock market during the crises due to their characteristic of safe haven investment. On the contrary, Low, Yao, and Faff (2016) affirms that precious metals give greater benefit than diamonds, which should be of one carat and above to be able considered as investment vehicle. Table 1 summarizes the main contributions reviewed.

The objective of this paper consists in analysing, from different perspectives, the determinants of diamond price, understanding their impact on price, and the marginal ratio of substitution between them. The knowledge of these relations will improve market efficiency, helping sellers in the price setting and buyers in the recognition of investment opportunities. In some cases, the need to sell and the location or the idiosyncrasy of the offer can make the price offered by the seller different from the one that could be expected from an econometric perspective, appearing then market opportunities that should be object of the buyers' attention.

Based on previous studies, our research conducts a multiple regression analysis of diamond price using a dataset of 166,026 diamonds that has been extracted from the online platform Blue Nile in July 2016. As Wolff (2015), we use a wide range of explanatory variables, which include not

[4]The dot-com crisis (2000) and the financial crisis (2007–2008).

only the classic 4Cs (carat weight, colour, clarity, and cut), but also the polishing, symmetry, and fluorescence attributes; all the diamonds in our sample are certified by the GIA. The econometric model controls price dispersion and allows comparing, for each diamond, the actual certified price and the predicted price according to the model (in-sample prediction).

At this point, it is important to emphasize that other techniques can be used to analyse and estimate the diamond price, apart from the regression technique. For instance, Wolff (2016) tries to measure the market efficiency of diamonds by using nonparametric data envelopment analysis instead of regression analysis. Holden and Serearuno (2002) publish an article in which they propose an artificial intelligence method to classify precious stones, including diamonds. Mosley (2013) develops a supervised machine learning model on a data set of 53,490 diamonds in order to perform a classification task. More recently, Wang (2016) uses statistical learning to design a digital gemmological analyser, whereas Şimşek and Buckmann (2017) focus on different types of heuristics to address the classification task. Finally, we would like to remark that other empirical strategies, such as regression trees and neural networks, may be suitable to analyse a large diamond sample like ours. Nonetheless, in our opinion, regression tree models are a simple predictive tool that does not allow hedonic analysis (sensitivity analysis). Meanwhile, neural networks require that all categorical variables be transformed into a set of binary properties, reaching a high complexity to the decision structure of this artificial intelligence method in large samples.

In spite of the different methodological approaches, the hedonic regression model is postulated in the literature as an adequate technique to analyse the formation of the diamond price—see, for example, Chu (2001) and Wolff (2016). Our regression model lets the comparison between the actual certified diamond price and the price predicted by the model. This comparison allows us to identify the "efficiency" grade (or potential profitability for the buyer) of each diamond. The measure of efficiency is oriented, in our study, towards the final customer, not towards the supplying companies, so that it is defined as efficient the stone that presents a lower price than the one expected according to the econometric model. Under this methodology, the best bargains in the market should be the target of potential buyers or investors. Other contributions of our regression model are the identification and analysis of focal points, the inclusion in the model of significant interactions between explanatory variables, and the definition of indifference curves (weight–quality, weight–colour, and weight–cut) for different budget levels.

The rest of the paper is organized as follows: after this introduction, Section 2 offers a general description of our parametric methodological approach. Section 4 makes a description of the sample of 166,026 diamonds; all units were on sale at Bluenile.com in July 2016. Section 4 applies the proposed methodology to the sample; estimated results and main findings are shown in this section. Finally, Section 5 summarizes the main conclusions of our research.

2 | METHODOLOGICAL APPROACH

The empirical analysis carried out in the next section tries to estimate two relations: (a) to what extent the different characteristics of a diamond determine its price and (b) how, given a certain price of the diamond, its weight is related to other characteristics. The following simplified linear regression relates the price of each diamond i to a constant term α_0, the carat weight of the diamond, a set of dummy variables $\{D_0, ..., D_k, ..., D_n\}$ that allow to identify each particular category $\{c_0, ..., c_k, ..., c_n\}$ of the explanatory categorical variable C (for example, C can represent the diamond colour), and, finally, an idiosyncratic error term ?.

$$\ln(price_i) = \alpha_0 + \beta \ln(carat_i) + \gamma_1 D_{1i} + ... + \gamma_k D_{ki} + ... + \gamma_n D_{ni} + \varepsilon_i \quad . \tag{1}$$

As price and carat weight are expressed in natural logarithms, we obtain a nonlinear relationship between those variables, representing the β coefficient the price elasticity of the diamond to its weight.

$$\beta = e_{price-carat} = \frac{\partial \ln(price_i)}{\partial \ln(carat_i)} = \frac{\Delta(\%)\, price}{\Delta(\%)\, carat} \quad . \tag{2}$$

As for the meaning of the dummy variables, we suppose that the category c_0 is the one that remains in the constant term of the regression. The percentage change on the price because of moving from the base category c_0 to the category c_k, with the other factors of the equation being equal, is obtained as follows:

$$\left.\begin{array}{l}\ln\left(price_{i;C_0}\right) = \alpha_0 + \beta \ln(carat_i) + 0 + \varepsilon_i \\ \ln\left(price_{i;C_k}\right) = \alpha_0 + \beta \ln(carat_i) + \gamma_k + \varepsilon_i\end{array}\right\} \Rightarrow \ln\left(price_{i;C_k}\right) - \ln\left(price_{i;C_0}\right) = \ln\left(\frac{price_{i;C_k}}{price_{i;C_0}}\right) = \gamma_k \Rightarrow$$

$$\Rightarrow \frac{price_{i;C_k}}{price_{i;C_0}} = e^{\gamma_k} \Rightarrow \frac{price_{i;C_k} - price_{i;C_0}}{price_{i;C_0}} = e^{\gamma_k} - 1.$$

Our sample will allow us to obtain an estimate of the coefficients studied: $\{\widehat{\alpha}_0, \widehat{\beta}, \widehat{\gamma}_1, ..., \widehat{\gamma}_k, ..., \widehat{\gamma}_n\}$. In the estimated equation, the residual $\widehat{\varepsilon}_i$ relates the actual price of the diamond and its predicted price as follows:

TABLE 2 Sample characteristics; continuous variables

Variable	Mean	*SD*	Min	Max
Carat weight	0.88	0.74	0.23	21.70
Price (€)	7,611	28,635	255	2,970,782
Price per carat	5,155	5,178	925	160,949

$$\ln(price_i) = \widehat{\alpha}_0 + \widehat{\beta}\ln(carat_i) + \widehat{\gamma}_1 D_{1i} + \ldots + \widehat{\gamma}_k D_{ki} + \ldots + \widehat{\gamma}_n D_{ni} + \widehat{\varepsilon}_i = \ln(\widehat{price_i}) + \widehat{\varepsilon}_i \Rightarrow$$
$$\Rightarrow \ln(price_i) - \ln(\widehat{price_i}) = \widehat{\varepsilon}_i \Rightarrow \ln\left(\frac{price_i}{\widehat{price_i}}\right) = \widehat{\varepsilon}_i \Rightarrow \frac{price_i - \widehat{price_i}}{\widehat{price_i}} = e^{\widehat{\varepsilon}_i} - 1 .$$

The second analysis we propose consists of estimating the marginal rate of substitution between the weight of the diamond and the rest of characteristics given a certain price level $\overline{P}$. If we order numerically the categorical variable C from the worst to the best category (in terms of quality), we can estimate the following equation:

$$\ln\left(carat_{i;\overline{P}}\right) = \varphi_0 + \delta C_i + \ldots + \varepsilon_i . \quad (3)$$

As carat weight is expressed in natural logarithms, the δ coefficient represents the semielasticity of the carat weight to the attribute C.

$$\delta = semi-elasticity_{carat-C} = \frac{\partial \ln(carat_i)}{\partial C} = \frac{\Delta(\%)\, carat}{\Delta C} . \quad (4)$$

3 | DESCRIPTION OF THE DATA

The analysed sample contains information, for a total amount of 166,026 diamonds,[5] on the sale price and on the main characteristics of each unit: carat weight, shape, cut, colour, and clarity. Table 2 shows descriptive statistics of the continuous variables: price[6] (€), carat weight, and price per carat. The average price and the average price per carat are, respectively, 7,611 € and 5,155 €; both distributions show a high dispersion around their means, especially the one of the price.

Table 3 contains the frequency observed when crossing the different shape categories with the different categories of the variables clarity, colour, and cut; the darker the cell, the higher the observed frequency. The most frequent shape is the round one, followed by the princess shape. Within the round shape, the most observed diamonds are those with intermediate clarity (from SI1 to VS2), D, E, and F colours, and ideal cut.

Figure 1 shows the histogram of the weights. We represent weights equal or less than three carats, which represent more than 97% of the sample. The carat weight is measured in tenths and hundredths, being more frequent the weights with a single decimal: 0.5 (7.34% of the sample), 0.4 (5.73%), 0.7 (5.57%), 1.01 (5.14%), 0.3 (4.86%), 1 (3.14%), and so forth.

Figure 2 illustrates two scatter plots: (a) the relation between carat weight and price and (b) the relation between carat weight and the mean price for each carat weight. The diamond price tends to increase with its weight in a relation that appears to be nonlinear. In addition, the dispersion of prices is increasing significantly with the carat weight. Regarding the relation between carat weight and the mean price for each carat weight, it is observed that the higher the weight, the higher the mean price, which occurs because the diamond prices grow (on average) more than proportionally with the carat weight. Moreover, the mean price presents discontinuities in certain weight levels; for example, it experiences a significant jump when the weight reaches 0.5 carats or when it reaches 1 carat and it falls when the weight approaches 0.7 carats. This fact seems to indicate that the market presents a higher degree of activity at certain weight levels or “focal points.”[7]

Figure 3 contains the mean “price per carat weight” observed for the different characteristics of diamonds. The three graphs in the figure differentiate between round diamonds (73.6% of the sample) and the other shapes (26.4%). The price of the diamond goes down as it worsens some of its characteristics, but this fall is more pronounced in the first levels of quality and it is especially observed in the case of clarity. It is also noted that round diamonds are generally cheaper than the other shapes at all levels of the different attributes except at the highest levels of clarity (FL) and cut (Signature Ideal).

[5]Other authors such as Scott and Yelowitz (2010), Lee et al. (2014), and Wolff (2015) analyse samples of similar size for the years 2005, 2012, and 2013, respectively; their data are coming from different sources. Our sample comes entirely from the same digital supplier (Blue Nile) and, therefore, it does not suffer distortions associated with the commissions of the sellers.

[6]Blue Nile group sets diamond prices based on the offers received from diamond owners (companies and individual investors) to which it adds its cost and trade margin; therefore, the diamond supply chain and the hedonic specifications of each diamond are not determinants in the Blue Nile pricing mechanism.

[7]This phenomenon is analysed by Wolff (2015), but he focuses on price volatility in general rather than on the significant price jumps when we approach or reach the focal points.

TABLE 3 Frequencies; categorical attributes

	Clarity								Colour							Cut				
Shape	FL	IF	VVS1	VVS2	VS1	VS2	SI1	SI2	D	E	F	G	H	I	J	Good	Very Good	Ideal	Siganture Ideal	Total
Round	227	5,794	12,832	18,465	23,435	25,776	21,850	13,789	22,880	22,188	21,186	19,462	15,897	11,941	8,614	9,015	27,799	83,255	2,099	122,168
Princess	41	445	1,761	1,827	2,695	2,799	2,231	904	1,522	2,922	2,655	2,221	1,512	1,098	773	3,321	8,000	0	1,382	12,703
Pear	16	642	1,159	987	1,130	1,285	1,173	670	1,984	1,397	1,254	1,139	613	458	217	913	6,149	0	0	7,062
Emerald	38	411	1,195	1,107	1,301	1,086	726	306	1,075	1,260	1,137	1,023	766	539	370	1,079	4,750	0	341	6,170
Cushion	6	123	413	584	1,077	1,391	1,270	430	629	889	1,021	978	784	582	411	1,806	3,383	0	105	5,294
Oval	3	265	606	587	852	1,008	986	492	1,211	938	774	752	457	390	277	840	3,959	0	0	4,799
Radiant	7	76	251	184	404	553	485	183	255	402	432	444	291	183	136	423	1,720	0	0	2,143
Asscher	23	123	423	372	524	432	144	34	297	464	453	447	222	138	54	355	1,193	0	527	2,075
Marquise	9	169	223	188	308	356	411	276	612	343	303	268	163	157	94	459	1,481	0	0	1,940
Heart	14	123	167	177	252	307	423	209	286	289	369	301	231	149	47	259	1,413	0	0	1,672
Total	384	8,171	19,030	24,478	31,978	34,993	29,699	17,293	30,751	31,092	29,584	27,035	20,936	15,635	10,993	18,470	59,847	83,255	4,454	166,026

Note. Clarity grades I1, I2, and I3 do not appear in our sample.

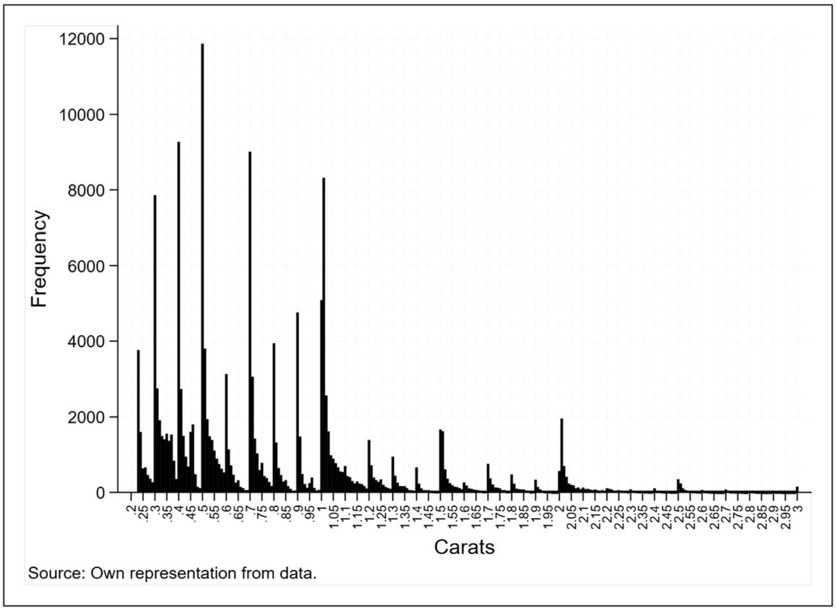

FIGURE 1 Histogram of variable "carat weight." Carat weights less than 3

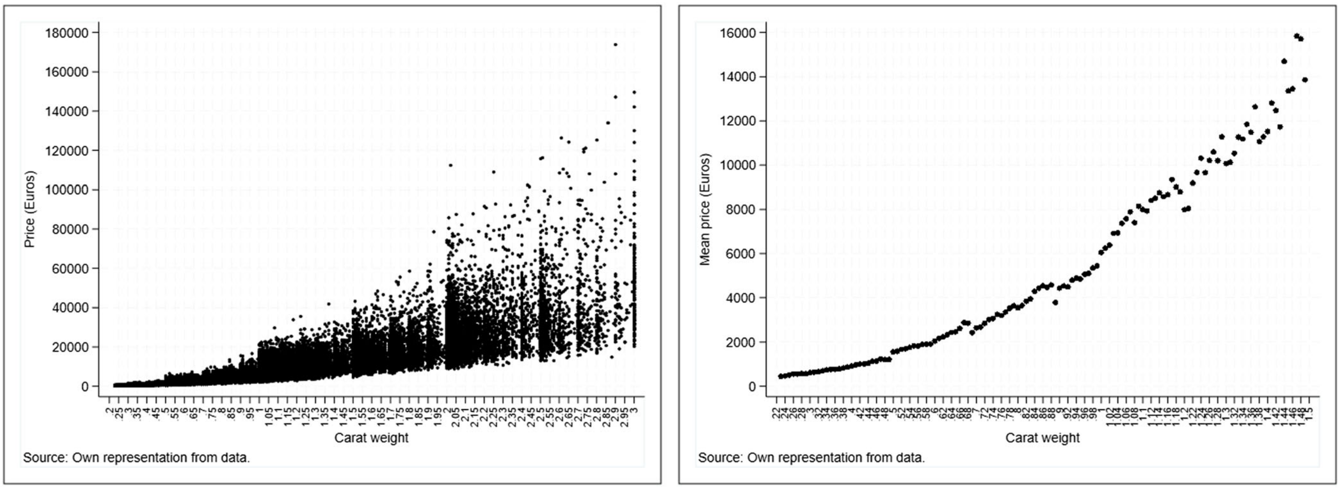

FIGURE 2 The relation between carat weight and "price" and "mean price for each carat weight"

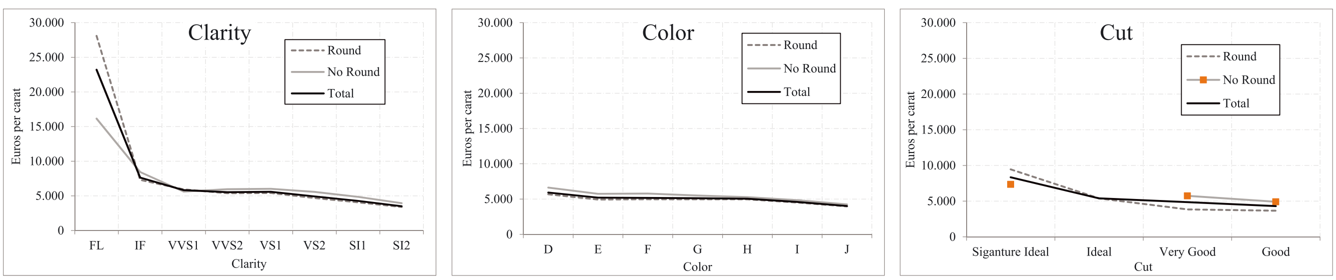

FIGURE 3 Mean "price per carat weight" by attributes

Our data description finishes with the analysis of the frequency (number of diamonds) and the mean price for each carat weight as Figure 4 represents. The aforementioned figure also shows the percentage change in the mean price for each carat weight when we move from a certain weight to the weight immediately following. For a clearer view of the graphs, Figure 4 depicts only diamonds less than 1.02 carats, which represent more than 75% of the sample. The diamond distribution is clearly concentrated in "weight focal points," which are given by whole weight values or slightly higher values than the whole ones,[8] note that there are considerably fewer diamonds available for sale on the left side of a focal point than on the right side. As in other studies—see, for example, Scott and Yelowitz (2010)—we find significant jumps in prices at focal points.[9]

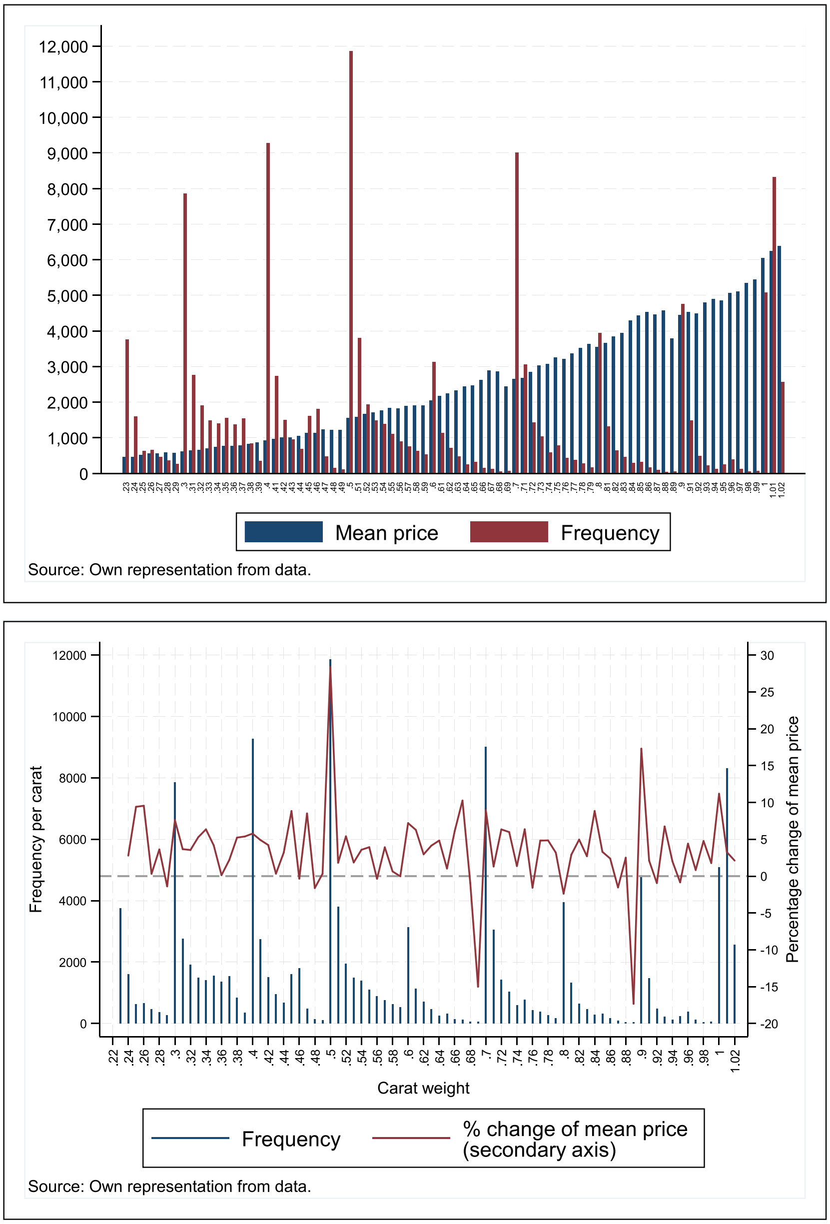

FIGURE 4 Frequency and mean price (level and percentage change) for each carat weight

For example, the price grows almost 30% when it goes from 0.49 to 0.5 carats. However, there are also significant price declines just before reaching some focal points; for instance, the price falls more than 15% when it goes from 0.88 to 0.89 carats. As we mentioned before, in these values close to a focal point (on the left side), the number of offered diamonds is relatively small.

4 | EMPIRICAL ANALYSIS

Table 4 contains a total of seven multiple linear regression estimates that try to explain the main determinants of the diamond prices.[10] The first three estimates focus on diamonds whose weight ranges from the lowest observed value (0.23 carats) to 4.1 carats (we cover more than 99% of the sample).[11] The four following estimates analyse diamonds whose weight corresponds to the most important focal points and their close values

[8]The sample has 5,084 diamonds of 1.00 carat and 8,315 diamonds of 1.01 carat, which indicates that diamond cutters prefer to add a little more weight than to stay short.

[9]Scott and Yelowitz (2010) analysed this effect only for round diamonds, while a broader analysis is made in this paper, reaching more specific conclusions.

[10]Other authors developed an empirical analysis on the diamond market which is based in other methodologies, as Wolff (2016), using Data Envelopment Analysis. On the other hand, Renneboog and Spaenjers (2012), Wolff (2015) and Lee et al. (2014) use regression analysis too, but with fewer explanatory variables and for the full price range, not analysing in detail possible focal points.

[11]Carat weights greater than 4.1 have not been considered because of their low frequency and high price dispersion.

[12]Full results are available to any reader who requests them.

TABLE 4 The determinants of diamond price

	Basic model	4C's model	Full model	(0.5–0.55) carats	(1–1.05) carats	(1.5–1.55) carats	(2–2.05) carats
Ln (Carat)	1.9 *0.00*	2 *0.00*	2.01 *0.00*				
Oval		−0.148 *0.00*	−0.131 *0.00*	−0.213 *0.00*	−0.165 *0.00*	−0.030 *0.18*	−0.104 *0.00*
Pear		−0.156 *0.00*	−0.139 *0.00*	−0.197 *0.00*	−0.148 *0.00*	−0.058 *0.03*	−0.104 *0.00*
Marquise		−0.173 *0.00*	−0.148 *0.00*	−0.173 *0.00*	−0.165 *0.00*	−0.077 *0.05*	−0.139 *0.00*
Princess		−0.205 *0.00*	−0.173 *0.00*	−0.244 *0.00*	−0.252 *0.00*	−0.104 *0.00*	−0.148 *0.00*
Heart		−0.156 *0.00*	−0.189 *0.00*	−0.181 *0.00*	−0.213 *0.00*	−0.213 *0.00*	−0.197 *0.00*
Cushion		−0.213 *0.00*	−0.197 *0.00*	−0.267 *0.00*	−0.267 *0.00*	−0.077 *0.00*	−0.156 *0.00*
Radiant		−0.221 *0.00*	−0.197 *0.00*	−0.288 *0.00*	−0.252 *0.00*	−0.104 *0.00*	−0.139 *0.00*
Asscher		−0.244 *0.00*	−0.229 *0.00*	−0.281 *0.00*	−0.274 *0.00*	−0.148 *0.00*	−0.229 *0.00*
Emerald		−0.274 *0.00*	−0.252 *0.00*	−0.336 *0.00*	−0.281 *0.00*	−0.156 *0.00*	−0.197 *0.00*
IF		−0.252 *0.00*	−0.244 *0.00*	−0.213 *0.00*	−0.213 *0.00*	−0.113 *0.11*	−0.068 *0.17*
VVS1		−0.343 *0.00*	−0.330 *0.00*	−0.330 *0.00*	−0.316 *0.00*	−0.244 *0.00*	−0.205 *0.00*
VVS2		−0.387 *0.00*	−0.381 *0.00*	−0.393 *0.00*	−0.375 *0.00*	−0.295 *0.00*	−0.244 *0.00*
VS1		−0.423 *0.00*	−0.417 *0.00*	−0.429 *0.00*	−0.411 *0.00*	−0.330 *0.00*	−0.302 *0.00*
VS2		−0.462 *0.00*	−0.451 *0.00*	−0.462 *0.00*	−0.457 *0.00*	−0.375 *0.00*	−0.375 *0.00*
SI1		−0.532 *0.00*	−0.523 *0.00*	−0.537 *0.00*	−0.523 *0.00*	−0.446 *0.00*	−0.451 *0.00*
SI2		−0.593 *0.00*	−0.585 *0.00*	−0.597 *0.00*	−0.589 *0.00*	−0.537 *0.00*	−0.546 *0.00*
E		−0.068 *0.00*	−0.077 *0.00*	−0.077 *0.00*	−0.095 *0.00*	−0.077 *0.00*	−0.104 *0.00*
F		−0.104 *0.00*	−0.113 *0.00*	−0.104 *0.00*	−0.148 *0.00*	−0.148 *0.00*	−0.181 *0.00*
G		−0.173 *0.00*	−0.181 *0.00*	−0.139 *0.00*	−0.221 *0.00*	−0.213 *0.00*	−0.259 *0.00*
H		−0.237 *0.00*	−0.244 *0.00*	−0.181 *0.00*	−0.295 *0.00*	−0.309 *0.00*	−0.362 *0.00*
I		−0.349 *0.00*	−0.356 *0.00*	−0.295 *0.00*	−0.381 *0.00*	−0.405 *0.00*	−0.462 *0.00*
J		−0.446 *0.00*	−0.446 *0.00*	−0.381 *0.00*	−0.467 *0.00*	−0.503 *0.00*	−0.542 *0.00*
Ideal		−0.213 *0.00*	−0.197 *0.00*	−0.181 *0.00*	−0.156 *0.00*	−0.181 *0.00*	−0.165 *0.00*
Very good		−0.267 *0.00*	−0.252 *0.00*	−0.244 *0.00*	−0.237 *0.00*	−0.229 *0.00*	−0.213 *0.00*

(Continues)

TABLE 4 (Continued)

	Basic model	4C's model	Full model	(0.5–0.55) carats	(1–1.05) carats	(1.5–1.55) carats	(2–2.05) carats
Good		−0.302 *0.00*	−0.281 *0.00*	−0.309 *0.00*	−0.281 *0.00*	−0.259 *0.00*	−0.237 *0.00*
Dummy focal point	0.01 *0.00*	0.06 *0.00*	0.07 *0.00*				
Dummy strong negative % variation	−0.03 *0.03*	−0.08 *0.00*	−0.09 *0.00*				
Dummy strong positive % variation	0.01 *0.03*	−0.03 *0.00*	−0.03 *0.00*				
Constant	8.62 *0.00*	9.45 *0.00*	11.03 *0.00*	8.46 *0.00*	10.64 *0.00*	12.99 *0.00*	13.28 *0.00*
Number of observations	164,783	164,783	164,783	20,469	19,450	4,685	4,058
Log likelihood	−33,961.1	81,738.2	88,058.1	16,752.6	12,901.7	3,380.3	2,702.5
Residual sum of squares	14,569.7	3,577.6	3,313.4	233.2	302.2	64.8	62.7
R-squared	94.0%	98.0%	99.0%	83.0%	87.0%	87.0%	88.0%
Adjusted R-squared	94.0%	98.0%	99.0%	83.0%	87.0%	87.0%	88.0%
AIC criterion	67,932.2	−163,416.3	−176,020.2	−33,425.1	−25,719.3	−6,680.6	−5,323.0
BIC criterion	67,982.3	−163,116.0	−175,539.6	−33,108.0	−25,388.5	−6,422.5	−5,064.4

Note. Table shows "exp (estimated coefficient) − 1" and *p* value.

Abbreviations: AIC, Akaike information criteria; BIC, Bayesian information criteria.

TABLE 5 Twenty investment opportunities in the diamond market

Diamond	Actual price	Estimated price	Actual/estimated	Carat	Shape	Clarity	Colour	Cut
Nonround shapes								
1	22,880	46,071	0.50	2.5	MQ	FL	H	Good
2	1,070	2,046	0.52	0.67	MQ	SI2	D	Very good
3	13,378	24,787	0.54	2.04	PS	FL	J	Very good
4	1,479	2,719	0.54	0.58	AS	FL	G	Very good
5	9,226	16,814	0.55	1.27	PR	FL	H	Signature ideal
6	30,433	54,457	0.56	2.8	EC	FL	H	Very good
7	1,663	2,942	0.57	0.57	RA	FL	E	Very good
8	2,205	3,885	0.57	0.71	EC	FL	F	Good
9	3,551	6,254	0.57	0.9	EC	FL	G	Very good
10	1,573	2,751	0.57	0.6	EC	FL	G	Very good
Round shape								
11	32,781	57,090	0.57	3.74	RD	SI2	G	Ideal
12	35,834	60,568	0.59	4.04	RD	SI2	I	Ideal
13	36,042	60,649	0.59	3.55	RD	VVS2	J	Ideal
14	10,889	17,749	0.61	1.52	RD	FL	J	Ideal
15	32,655	53,202	0.61	3.04	RD	IF	J	Ideal
16	49,108	79,407	0.62	3.5	RD	VVS1	H	Ideal
17	20,698	33,308	0.62	2.61	RD	VVS2	J	Ideal
18	26,137	42,039	0.62	3	RD	SI2	H	Ideal
19	24,086	38,730	0.62	3.01	RD	SI2	H	Ideal
20	37,326	59,880	0.62	4.01	RD	SI2	H	Ideal

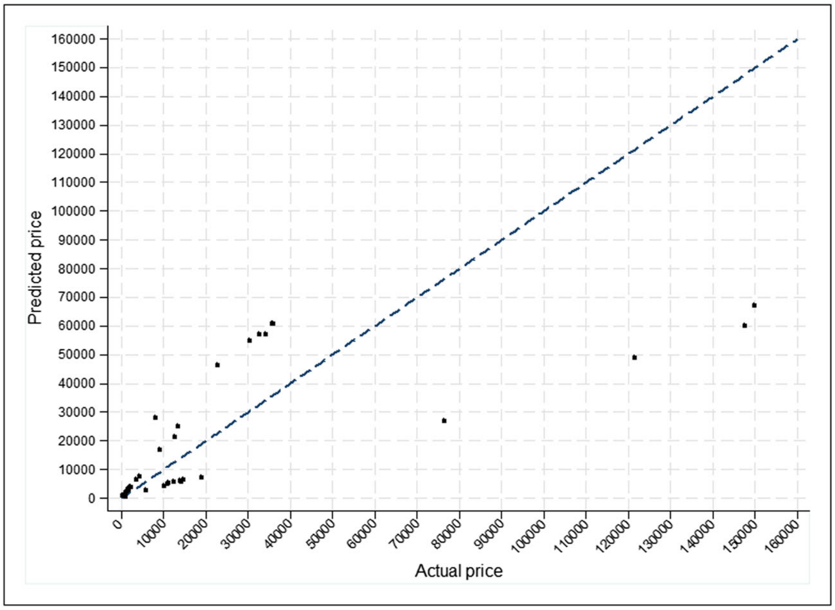

FIGURE 5 Largest differences between actual and predicted prices in the full model

by the right; that is, 0.5–0.55, 1–1.05, 1.5–1.55, and 2–2.05 carats; these four intervals, taken together, represent approximately 5% of the weight range (2.3, 4.1) and 30% of the sample. These four estimates will allow us to check if the market shows a different behaviour at different focal points. In all estimates, price and carat weight are expressed in natural logarithms; this log transformation allows to control for the non-linear relationship between the two mentioned variables and avoid heteroskedasticity problems associated to the larger price dispersion observed in higher levels of variable weight. The first estimate (basic model) regresses log-price on log-carat weight and on three dummy variables: one that controls focal points and two others that control relatively large percentage increases/decreases (when we move from one carat weight to the next one) of the diamond mean price for each carat weight; in particular, we mark the growths belonging to the highest/lowest deciles of the percentage changes distribution. The second model (4C's model) incorporates the previous estimation information on the shape of the diamond and on the other main characteristics clarity, colour, and cut. The omitted dummy variable categories in this estimate (and in the following ones) are "round shape," "D colour," "flawless clarity," and "signature ideal cut." The third model (full model) incorporates the second model additional features of diamonds that are, according to experts, less important than the previous ones for determining the diamond price but which should also be taken into account; we refer to fluorescence, depth, polish, and symmetry. The estimated coefficients for the different categories of these last variables are not shown in the table to avoid their oversize.[12] Finally, the last four estimates use the same specification as the full model but delimiting the weight in the above-mentioned focal ranges.

Regarding the first three models, the first one (Basic model), which only includes as explanatory variables the logarithm of carat weight and the dummies for focal points and extreme growths, shows a high adjusted R^2 of 94%; this coefficient raises to 98% when the attributes of shape, clarity, colour, and cut are incorporated into the specification in the second model (4C's model), and finally reaches 99% when we incorporate the rest of attributes into the third model (full model): fluorescence, depth, polish, and symmetry. The dummy variables controlling for focal points and abnormal growths are significant in the three models. The full model is the one that shows the smallest residual sum of squares and the smallest Akaike and Bayesian information criteria (on information criteria, see Sakamoto, Ishiguro, & Kitagawa, 1986), so we will take this latest model as the one that produces a better fit.

According to the full model, the elasticity of the price to the weight is Approximately 2; that is, if the carat weight of the diamond increases by 1%, its price increases by 2%. As for the shape of the diamond, the price falls more than 10% when the shape changes from round to any other, and this fall is especially sharp when we move from round to asscher or emerald, as it exceeds 20% in absolute value. Regarding clarity, it is observed that the fall in price is very pronounced (−24.4%) when we move from the best clarity (Flawless) to the following scale (Internally Flawless); from there, the successive jumps towards worse clarities cause more moderate price falls that do not exceed 10% in absolute value. As for colours, the largest price falls are observed at the beginning and at the end of the colour scale; that is, when we jump from D to E (−7.7%), from H to I (−11.2%), and from I to J (−9%). Finally, the price experiences a sharp fall of almost −20% when the cut changes from Signature Ideal to Ideal, and then it experiences more modest declines as we consider worse cuts.

As we mentioned above, the last four estimates of Table 4 use the same specification as the one of the full model but delimiting the carat weight in different focal ranges as follows: 0.5–0.55, 1–1.05, 1.5–1.55, and 2–2.05 carats. The main conclusions that can be drawn from these estimates are as follows: the drop in prices because of moving from the round shape to any other shape is significantly more moderate in the higher weight ranges, and

[13]Of these 25 diamonds, the top five have, each one, a potential surplus of more than 80%, and the best of them, which costs 8,133 Euros, has a potential return of 242.5%.

TABLE 6 The determinants of the price of round diamonds; models with interactions

cl-co model		cl-cu model		co-cu model	
Ln (Carat)	1.975***	Ln (Carat)	1.981***	Ln (Carat)	1.981***
IF	−0.315***	IF	−0.333***	IF	−0.415***
SI1	−0.587***	SI1	−0.476***	SI1	−0.528***
SI2	−0.717***	SI2	−0.572***	SI2	−0.618***
VS1	−0.823***	VS1	−0.661***	VS1	−0.677***
VS2	−0.895***	VS2	−0.741***	VS2	−0.744***
VVS1	−1.070***	VVS1	−0.875***	VVS1	−0.887***
VVS2	−1.213***	VVS2	−0.802***	VVS2	−1.026***
E	−0.347***	E	−0.064***	E	−0.103***
F	−0.473***	F	−0.100***	F	−0.170***
G	−0.726***	G	−0.174***	G	−0.281***
H	−0.814***	H	−0.254***	H	−0.393***
I	−1.006***	I	−0.420***	I	−0.549***
J	−1.284***	J	−0.573***	J	−0.531***
Signature ideal	−0.252***	Signature ideal	−0.211***	Signature ideal	−0.320***
Ideal	−0.320***	Ideal	−0.512***	Ideal	−0.427***
Very good	−0.400***	Very good	−0.588***	Very good	−0.508***
IF # E	0.100***	IF # Ideal	−0.092**	E # Ideal	0.031***
IF # F	0.141***	IF # Very good	0.12	E # Very good	0.059***
IF # G	0.244***	IF # Good	0.078	E # Good	0.062***
IF # H	0.208***	VVS1 # Ideal	−0.057	F # Ideal	0.060***
IF # I	0.237*	VVS1 # Very good	0.136	F # Very good	0.094***
IF # J	0.378***	VVS1 # Good	0.122	F # Good	0.093***
VVS1 # E	0.276***	VVS2 # Ideal	−0.054	G # Ideal	0.096***
VVS1 # F	0.340***	VVS2 # Very good	0.159	G # Very good	0.140***
VVS1 # G	0.465***	VVS2 # Good	0.156	G # Good	0.137***
VVS1 # H	0.437***	VS1 # Ideal	−0.03	H # Ideal	0.124***
VVS1 # I	0.437***	VS1 # Very good	0.199	H # Very good	0.183***
VVS1 # J	0.561***	VS1 # Good	0.196	H # Good	0.176***
VVS2 # E	0.265***	VS2 # Ideal	−0.02	I # Ideal	0.111***
VVS2 # F	0.353***	VS2 # Very good	0.221	I # Very good	0.173***
VVS2 # G	0.534***	VS2 # Good	0.224	I # Good	0.168***
VVS2 # H	0.517***	SI1 # Ideal	−0.029	J # Ideal	−0.061***
VVS2 # I	0.515***	SI1 # Very good	0.21	J # Very good	0.003
VVS2 # J	0.626***	SI1 # Good	0.204	J # Good	(omitted)
VS1 # E	0.309***	SI2 # Ideal	−0.244*		
VS1 # F	0.407***	SI2 # Very good	(omitted)		
VS1 # G	0.586***	SI2 # Good	(omitted)		
VS1 # H	0.586***				
VS1 # I	0.581***				
VS1 # J	0.694***				
VS2 # E	0.299***				
VS2 # F	0.391***				
VS2 # G	0.586***				

(Continues)

TABLE 6 (Continued)

cl-co model		cl-cu model		co-cu model	
Ln (Carat)	**1.975*** **	**Ln (Carat)**	**1.981*** **	**Ln (Carat)**	**1.981*** **
VS2 # H	0.612***				
VS2 # I	0.621***				
VS2 # J	0.738***				
SI1 # E	0.315***				
SI1 # F	0.415***				
SI1 # G	0.616***				
SI1 # H	0.647***				
SI1 # I	0.707***				
SI1 # J	0.837***				
SI2 # E	0.296***				
SI2 # F	0.392***				
SI2 # G	0.595***				
SI2 # H	0.646***				
SI2 # I	0.736***				
SI2 # J	0.927***				
Dummy focal point	0.090***	Dummy focal point	0.090***	Dummy focal point	0.088***
Dummy negative % var.	−0.101***	Dummy negative % var.	−0.102***	Dummy negative % var.	−0.100***
Dummy positive % var.	−0.030***	Dummy positive % var.	−0.028***	Dummy positive % var.	−0.026***
Constant	10.005***	Constant	9.853***	Constant	9.958***
Number of observations	121,569		121,569		121,569
Log likelihood	70,939.5		64,332.2		64,498.5
Residual sum of squares	2,215.6		2,470.1		2,463.3
R-squared	99.0%		99.0%		99.0%
Adjusted R-squared	99.0%		99.0%		99.0%
AIC criterion	−141,753.0		−128,584.4		−128,921.1
BIC criterion	−141,141.4		−128,196.1		−128,552.2

Note. Table shows estimated coefficients.

Abbreviations: AIC, Akaike information criteria; BIC, Bayesian information criteria.

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

especially in the range of 1.5–1.55 carats; ergo, the effect on price of the diamond shape is less relevant (in absolute value) when considering heavier diamonds. Something similar happens with clarity and cut: the negative effect on the price of worsening either clarity or cut is smaller when we consider heavier diamonds, especially in the first (better) clarity levels and in the last (worse) cut levels. Only in the case of colours the opposite phenomenon is observed: colour changes appear to have a greater effect (in absolute terms) on the price of heavier diamonds.

The analysis of Table 4 concludes with the interpretation of the residuals. In each model, the exponential of the residual for each diamond indicates the ratio between the actual price of the diamond and its predicted value according to the model; ratios lower than one could indicate "profitable" diamonds or market opportunities, because these are diamonds whose predicted prices according to the model are higher than the ones actually observed (see the top 20 opportunities in Table 5); the opposite reasoning applies when the ratio is greater than one. Figure 5 shows some diamonds that have a relatively high or low price ratio. The diamonds below the bisector might be overvalued by the market; that is, their actual price is much higher than that assigned by the model according to their characteristics (price ratio >1); meanwhile, diamonds above the bisector have a current price lower than their expected value according to the model (price ratio <1), so that they could constitute a good investment. The 25 "unprofitable" diamonds with the highest ratio have an average weight of 2.2 carats, whereas the 25 profitable units[13] with the lowest ratio have an average weight of 1.4.

The next step in our analysis is to introduce into the model interactions between the explanatory variables. An interaction implies that the effect of one variable changes depending on the values (or categories) of some other variable. Table 6 shows estimates that include interactions

[14]We omit the *p* values to avoid too long a table.

TABLE 7 Marginal rate of substitution between carat weight and the rest of the main characteristics

	1,000	2,000	3,000	6,000	10,000	15,000
Colour	−0.03***	−0.04***	−0.04***	−0.04***	−0.06***	−0.07***
	0.000	0.000	0.000	0.000	0.000	0.000
Clarity	−0.04***	−0.05***	−0.05***	−0.04***	−0.06***	−0.06***
	0.000	0.000	0.000	0.000	0.000	0.000
Cut	−0.03***	−0.05***	−0.04***	−0.02***	−0.09***	−0.08***
	0.000	0.000	0.000	0.000	0.000	0.000
Constant	−0.53***	−0.05***	0.13***	0.33***	0.97***	1.14***
	0.000	0.000	0.000	0.000	0.000	0.000
Number of observations	3,925.0	2,060.0	1,030.0	500.0	197.0	81.0
Mean carat weight	0.41	0.58	0.72	1.01	1.22	1.58
Log likelihood	5,739.9	2,769.6	1,563.6	683.1	217.1	103.1
Residual sum of squares	12.3	8.2	2.9	1.9	1.3	0.4
R-squared	61.0%	77.0%	78.0%	70.0%	77.0%	91.0%
Adjusted R-squared	61.0%	77.0%	78.0%	70.0%	77.0%	91.0%
AIC criterion	−11,471.9	−5,531.2	−3,119.2	−1,358.2	−426.2	−198.1
BIC criterion	−11,446.8	−5,508.7	−3,099.5	−1,341.4	−413.1	−188.5

Abbreviations: AIC, Akaike information criteria; BIC, Bayesian information criteria.

[a]Table shows estimated coefficients (semi-elasticities). ***p<0.01

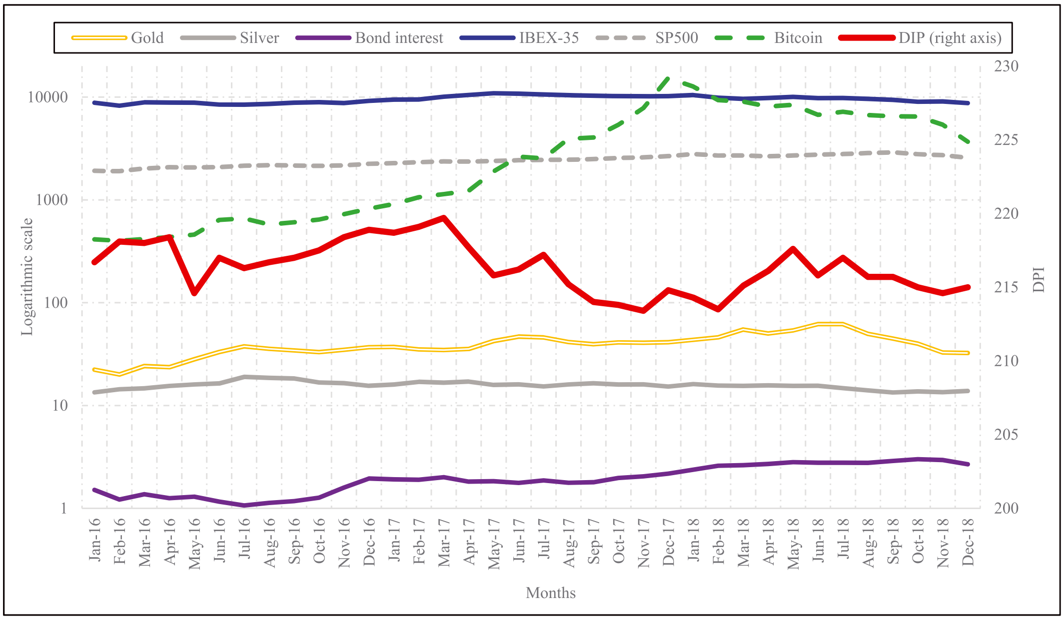

FIGURE 6 Comparative dynamic analysis of the diamond price in 2016–2018

of variables. In order to avoid too many interactions, we have analysed only round diamonds (74% of the sample) and proposed three different models[14]: one that models an interaction between clarity and colour (cl-co model), another that relates clarity with cut (cl-cu model), and a last one that relates colour with cut (co-cu model).

The results of the estimates reveal that the interaction of variables has in general significant effects on the price of the diamond; these effects can be positive or negative depending on the particular combination of categories. The first model (cl-co model) is the one that produces a better fit in terms of residual sum of squares and of information criteria. In this model, the negative coefficients of variables clarity and colour separately and the existence of some positive coefficients when they interact indicate that the price tends to fall as clarity or colour worsens, but that

[15]To ensure a minimum number of observations in all estimates, we have used narrow pricing intervals (rather than price levels), which are the results of subtracting and adding 50 euros to each reference price.

TABLE 8 ARIMA model for the DPI in the period 2016:01–2018:12

Endogenous variable: DPI (log)	Coef.	Std. Err.	z	*p* > *z*	(95% Conf. Interval)	
Gold price (log)	0.014	0.004	3.640	0.000	0.006	0.022
Silver price (log)	0.026	0.010	2.600	0.009	0.006	0.045
Bond interest (log)	0.039	0.005	7.870	0.000	0.029	0.048
SP500 index (log)	−0.062	0.017	−3.600	0.000	−0.096	−0.028
Bitcoin price (log)	−0.009	0.001	−8.790	0.000	−0.011	−0.007
Constant	5.783	0.121	47.990	0.000	5.547	6.019
AR structure						
L1.	−0.391	0.140	−2.810	0.005	−0.665	−0.118
L2.	−0.419	0.208	−2.020	0.044	−0.827	−0.012
L3.	−0.629	0.145	−4.340	0.000	−0.913	−0.345
σ_ε	0.004	0.001	5.730	0.000	0.002	0.005
Number of observations = 36		Wald chi2(8) = 654.76				
Log likelihood = 150.8001		Prob > chi2 = 0				

Note. Table shows estimated coefficients (elasticities).

Abbreviations: AR, autoregressive; ARIMA, autoregressive integrated moving average; DPI, Diamond Price Index.

downward trend is less severe in certain combinations of clarity and colour. For example, diamonds with medium–low clarity (from VVS2 to SI2) have lower prices than those with higher clarity, but within the diamonds of medium–low clarity, the units with colours from G to J experience a lower price fall than the units with better colours; this fact could be due to the greater relative demand of these combinations of diamonds in the market—many consumers looking for diamonds of medium–low clarity and an medium–low colour. In the other two models (cl-cu model and co-cu model), a relatively high positive effect is observed when Good or Very good cut categories interact with medium–low clarities or colours.

To come to an end, our empirical analysis, we will try to identify the marginal rate of substitution between diamond weight and the rest of main characteristics (colour, clarity, and cut). We will study this marginal rate by estimating the relation between the carat weight (of round diamonds) and the other characteristics when the price of the diamond remains fixed at a certain level. We will use different price levels to be able to contrast if marginal rates are more or less stable. The price levels that we are going to consider are 1,000, 2,000, 3,000, 6,000, 10,000, and 15,000 euros.[15] Table 7 shows the estimated coefficients:

As the carat weight is expressed in logarithms and the explanatory variables have been ordered numerically from the worst to the best category, the coefficients represent semielasticities of substitution. The estimations show coefficients with a negative sign (as might be expected) and always greater than −0.1. The highest coefficient (in absolute value) is found in the relation between weight and cut in diamonds of 10,000 or 15,000 euros: in these cases, to keep the price constant, if the diamond improves its cut one level, the carat weight should be 8%–9% lower. Another aspect to observe is that marginal rates of substitution tend to be higher when we consider higher prices, although this trend is not observed in the diamonds of 6,000 euros, possibly because of their proximity to the 1-carat focal point; in diamonds whose carat weight belongs to focal points (or focal intervals), the substitution rates are smaller because of the buyer's unwillingness to give up a certain weight.

To conclude our analysis, we have tried to obtain some evidence of the relationship between the diamond price and those external factors that could have some influence on that price, factors such as the business cycle, seasonal variations, and the opportunity cost of investing in diamonds (share/bond prices, gold prices, etc.). The static nature of our dataset (cross-sectional data of year 2016) does not allow a dynamic analysis—that is, external factors are considered fixed data in our study—however, we can use other databases to explore the proposed relationships. With this objective, we have obtained monthly time series, from January 2016 to December 2018, that allow us to relate a Diamond Prices Index (DPI)[16] with the value of other investment alternatives. Specifically, we analyse the temporal evolution of the following assets: DPI, S&P-500 Index, IBEX-35 Index, Bitcoin value in USD, interest rate of 5-year American bonds, and gold (100 oz.) and silver prices (in USD); all these series have been obtained from the URL finance.yahoo.com, except the DPI indicator that comes from www.diamondse.info.

As Figure 6 shows, the DPI indicator grows from mid-2016 until the spring of 2017, moment at which it begins to fall significantly until February 2018, as of this month, the index recovers from the previous fall but without reaching the levels of early 2017. To contrast the consistency of this

[16]The Diamond Prices Index is a representation of the current market pricing trend for diamonds. It is published by the company Fairfield County Diamonds Inc. on the website http://www.diamondse.info/diamonds-price-index.asp. The DPI takes into account the average retail price per carat of loose diamonds from jewellers around the web. Prices are calculated for groups of weight ranges as well as by colour and clarity.

[17]Autoregressive integrated moving average (ARIMA) fits univariate models for a time series, where the disturbances are allowed to follow a linear autoregressive moving-average specification. When independent variables are included in the specification, such models are often called ARMAX models. On ARIMA models see, for example, Hamilton (1994).

[18]The regression shows a positive relationship between the Diamond Price Index and the bond interest, which necessarily implies a negative relationship between the prices of both assets.

evolution, we have compiled a sample of 18,338 diamonds that have been offered on the Blue Nile website in July 2016, March 2017, and December 2018. The price per carat of the diamonds rose by 3.9% between July 2016 and March 2017, and fell by −9.7% between February 2017 and December 2018; this behaviour is comparable with the one described by the DPI index. The Blue Nile diamond with the highest price fall between 2016 and 2018 has been the one with the Asscher shape (−10.4%), whereas the one with the lowest fall has been the round diamond (−4.5%).

To quantify the relationships observed in Figure 6, we have estimated an autoregressive (AR) model of the DPI indicator as shown in Table 8. The variable DPI has turned out to be clearly stationary (applying the augmented Dickey–Fuller unit-root test), and follows, according to its autocorrelation structure, an AR scheme of order 3 (ARIMAX(3,0,0))[17]—this AR structure indicates that the endogenous variable shows a high inertia. The monthly dummies to control for seasonality have turned out to be non-significant, so they have not been included in the final version of the estimation. All the variables in the AR model (the explanatory ones and the explained one) have been expressed in logarithms to improve the model fit, so that the estimated coefficients can be interpreted as elasticities. The estimates show a positive (and inelastic) relationship between the diamond price index and the price of other commodities (gold or silver), and a negative (and inelastic) relationship between that index and the price of the financial assets (SP500, Bond price,[18] and Bitcoin). These observed elasticities suggest that real commodities, on the one hand, and financial assets, on the other hand, have a substitute character.

5 | CONCLUSIONS

Diamonds are not the typical economic merchandise. On the demand side, they are consumed not only for their intrinsic utility but also for the impression their consumption has on other people. On the supply side, the business of diamonds has traditionally presented two main drawbacks: the difficulty of knowing and assessing the physical characteristics of each unit and the complexity of a business where various market structures can be found (monopoly, oligopoly, and perfect competition).

We have collected data on diamond prices and other diamond attributes from a large online diamond supplier (Bluenile.com) with the purpose of analysing the determinants of the price by estimating different multiple linear regressions. We have used a wide range of explanatory variables, which include not only the classic 4Cs characteristics (carat weight, colour, clarity, and cut) but also the polishing, symmetry, and fluorescence attributes; all the diamonds in our sample are certified by the GIA. The econometric model controls heteroskedasticity (price dispersion) and allows comparing, for each of the 166,026 diamonds in the sample, the actual certified price and the predicted price according to the in-sample model prediction. Through this comparison, we can try to identify highly profitable units ("efficient" units) for buyers or investors.

Our estimates also shed light on the retail diamond market on the supply side. The weight of the diamond is the most determining variable in order to explain the price. According to our estimates, the elasticity of the price to the carat weight is approximately 2; that is, if the carat weight of the diamond increases by 1%, its price increases by 2%. As in previous studies, we have found significant jumps in prices at focal points (half- and whole-carat weights). For example, the price grows almost 30% when it goes from 0.49 to 0.5 carats. However, there are also significant price declines just before reaching some focal points; for instance, the price falls more than 15% when it goes from 0.88 to 0.89 carats, possibly due to the absence of a market in those carat weights. This fact has led us to estimate the relationship between price and its determinants considering also different focal ranges of weights: 0.5–0.55, 1–1.05, 1.5–1.55, and 2–2.05 carats. Taking into account these ranges, we observe that the drop in prices because of moving from the round shape to any other shape is significantly more moderate in the higher weight ranges. Something similar happens with clarity and cut: the negative effect on the price of worsening either clarity or cut is smaller when we consider heavier diamonds, especially in the first (better) clarity levels and in the last (worse) cut levels. Only in the case of colours the opposite phenomenon is observed: colour changes appear to have a greater effect (in absolute terms) on prices of heavier diamonds.

Another contribution of our study has been the inclusion, in a model on the price of round diamonds (the most frequent shape), of significant interactions between explanatory variables. For example, the negative coefficients of variables clarity and colour separately and the existence of some positive coefficients when they interact seem to indicate that price tends to fall as clarity or colour worsens, but that downward trend is interrupted in certain combinations of clarity and colour; this fact could be due to the greater relative demand of those combinations of diamonds in the market.

Our analysis concluded with the estimation of "indifference" curves, on the supply side, between the carat weight (the explained variable in these estimations) and the attributes clarity, cutting ,and colour. The estimates are conducted for round diamonds and for different price levels given. These estimates show coefficients with a negative sign, as might be expected. The highest coefficient (in absolute value) is found in the relation between weight and cutting in diamonds of 10,000 or 15,000 euros: in these cases, in order to keep the price constant, if the diamond improves its cut one level, the carat weight should be 8%–9% lower. Another aspect to observe is that marginal rates of substitution tend to be higher when we consider higher prices, although this trend is not observed in the diamonds of 6,000 euros, possibly because of their proximity to the 1-carat focal point; in diamonds whose carat weight belongs to focal points (or focal intervals), the substitution rates are smaller because of the buyer's unwillingness to give up a certain weight.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Fernando Núñez https://orcid.org/0000-0001-8029-2493

Antonio J. Tallón-Ballesteros https://orcid.org/0000-0002-9699-1894

REFERENCES

Auer, B. R. (2014). Could diamonds become an investor's best friend? *Review of Managerial Science*, *8*(3), 351–383. https://doi.org/10.1007/s11846-013-0108-2

Auer, B. R., & Schuhmacher, F. (2013). Diamonds—A precious new asset? *International Review of Financial Analysis*, *28*, 182–189. https://doi.org/10.1016/j.irfa.2013.03.008

Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, *45*(2), 217–229.

Bergenstock, D. J., Deily, M. E., & Taylor, L. W. (2006). *A cartel's response to cheating: An empirical investigation of the De Beers diamond empire. In The Global Diamond Industry* (pp. 95–120). UK: Palgrave Macmillan.

Chu, S. (2001). Pricing the C's of diamond stones. *Journal of Statistics Education*, *9*(2). https://doi.org/10.1080/10691898.2001.11910659

Ekelund, R. B. Jr., & Thornton, M. (2011). Galileo, Smith and the paradox of value: the 'connection' of art and science. *History of Economic Ideas*, 85–101.

Grynberg, R., & Mbayi, L. (Eds.) (2015). *The global diamond industry: Economics and development*, (Vol. 2), Springer.

Gupta, S., Polonsky, M., Woodside, A., & Webster, C. M. (2010). The impact of external forces on cartel network dynamics: Direct research in the diamond industry. *Industrial Marketing Management*, *39*(2), 202–210. https://doi.org/10.1016/j.indmarman.2008.11.009

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Holden, T., & Serearuno, M. (2002). Artificial intelligence for the design and grading of precious stones. In *Artificial Intelligence in Design'02* (pp. 237–258). Dordrecht: Springer.

Joy, M. (2011). Gold and the US-Dollar: Hedge or haven? *Finance Research Letters*, *8*, 120–131.

Lee, J. A., Caudill, S. B., & Mixon, F. G. Jr. (2014). Shine bright like a diamond: A hedonic model of grading and pricing an experience good. *Applied Economics*, *46*(16), 1829–1838. https://doi.org/10.1080/00036846.2014.884707

Levin, M., & Sheveleva, I. (2010). Institutional aspects of the organization and operation of the international diamond market. *Problems of Economic Transition*, *52*(10), 21–43. https://doi.org/10.2753/PET1061-1991521002

Low, R. K. Y., Yao, Y., & Faff, R. (2016). Diamonds vs. precious metals: What shines brightest in your investment portfolio? *International Review of Financial Analysis*, *43*, 1–14. https://doi.org/10.1016/j.irfa.2015.11.002

Mosley, Lawrence. A balanced approach to the multi-class imbalance problem. PhD Thesis, Iowa State University, Graduate Theses and Dissertations 13537, 2013.

Renneboog, L., & Spaenjers, C. (2012). Hard assets: The returns on rare diamonds and gems. *Finance Research Letters*, *9*(4), 220–230. https://doi.org/10.1016/j.frl.2012.07.003

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Tokyo: KTK Scientific Publishers.

Scott, F., & Yelowitz, A. (2010). Pricing anomalies in the market for diamonds: Evidence of conformist behavior. *Economic Inquiry*, *48*(2), 353–368. https://doi.org/10.1111/j.1465-7295.2009.00237.x

Şimşek, Özgür; Buckmann, Marcus. On Learning Decision Heuristics. In Imperfect decision makers: Admitting real-world rationality. Proceedings of Machine Learning Research, 2017. p. 75–85.

Smith, Adam. (1786). An inquiry into the nature and causes of the wealth of nations vol. 2.

Spar, D. L. M. (2006). Continuity and change in the international diamond market. *The Journal of Economic Perspectives*, *20*(3), 195–208. https://doi.org/10.1257/jep.20.3.195

Von Saldern, M. (1992). Forecasting rough diamond prices: A non-linear optimization model of dominant firm behaviour. *Resources Policy*, *18*(1), 45–58. https://doi.org/10.1016/0301-4207(92)90052-B

Wolff, F.-C. (2015). Does price dispersion increase with quality? Evidence from the online diamond market. *Applied Economics*, *47*(55), 5996–6009. https://doi.org/10.1080/00036846.2015.1061645

Wolff, F.-C. (2016). Bargaining powers of buyers and sellers on the online diamond market: a double perspective non-parametric analysis. *Annals of Operations Research*, *244*(2), 697–718. https://doi.org/10.1007/s10479-016-2160-1

AUTHOR BIOGRAPHIES

Angel Arcos Vargas is Professor at the School of Engineering of the University of Seville (Spain). He holds a PhD in Industrial Engineering from the University of Seville, a Diploma in Top Management (AD1) and in Top Management of Leading Companies (ADEL) from the San Telmo International Institute, and a Diploma in Leadership for Energy (L4E) from the Harvard Business School. His research focuses on the distribution business, its regulation, market structure, and impact of technological changes on efficiency. Renewables, energy storage, and electric vehicles, on which he has more than 50 publications and research projects. His professional career has been developed during more than 25 years in the Electric Sector, in multinational companies, where he occupied several positions of Top Direction. Within the electricity sector, his area of specialisation is distribution and network business. He is currently Chairman of the AENOR "Sistemas de suministro de energía

eléctrica" Technical Committee for Standardisation and Member of the Board of the Royal Academy of Engineering of Spain. He has belonged to the Management Committees of CIRED, CIGRE, CIDEL, FUTURED, and EDSO 4 Smartgrid, in which he was a founding member.

Fernando Núñez obtained BA (1997) and PhD (2006, with distinction) in Economics. Economics lecturer since 2001. Research stays, among other places, at Carlos III University (Spain) and the Universities of Manchester and Essex (both in the United Kingdom). Current academic secretary of the Department of Industrial Organization and Business Management I at the University of Seville (from 2013). His main research areas are in data mining, labor economics, and economy of the electricity sector.

Antonio J. Tallón-Ballesteros received the Technical Engineer degree in Systems Computer Science from the University of Córdoba, Spain, in 2002, the MSc degree in Computer Science from the University of Granada, Spain, in 2004, and the PhD degree in Computer Science from the University of Seville, Spain, in 2013. He also got the accreditation for Senior Lecturer in 2013. He is a PhD Lecturer at the University of Seville in the Department of Languages and Computer Systems. His current research areas include neural networks, evolutionary algorithms, machine learning, and data mining.

How to cite this article: Arcos-Vargas A, Núñez F, Tallón-Ballesteros AJ. Bargains in the diamond market? How to take advantage from online information. *Expert Systems*. 2019;e12423. https://doi.org/10.1111/exsy.12423