Financial Internet Quarterly „e-Finanse" 2016, vol.12 / nr 1, s. 21 - 31
DOI: 10.14636/1734-039X_12_1_003

eFinanse
financial internet quarterly

TOWARDS DATA-ORIENTED ANALYSIS OF THE ART MARKET: SURVEY AND OUTLOOK

Dominik Filipiak[1], Agata Filipowska[2]

Abstract

Due to the constantly growing interest in alternative investments, the art market has become the subject of numerous studies. By publishing sales data, many services and auction houses provide a foundation for further research on the latest trends. Determining the definition of the artistic value or formalisation of appraisal may be considered quite complex. Statistical analysis, econometric methods or data mining techniques could pave the way towards better understanding of the mechanisms occurring on the art market. The goal of this paper is to identify, describe and compare solutions (and related challenges) that help to analyse, make decisions and define state of the art in the context of the intersection of econometrics on art markets and computer science. This work is also a starting point for further research.

JEL classification: : C43, C51, G11, Z11
Keywords: art market analysis, hedonic regression

Received: 01.08.2015 Accepted: 14.03.2016

1 Department of Information Systems, Poznań University of Economics, e-mail: dominik.filipiak@kie.ue.poznan.pl.
2 Department of Information Systems, Poznań University of Economics, e-mail: agata.filipowska@kie.ue.poznan.pl.

Introduction

Plattner (1998) has written *about a market where producers don't make work primarily for sale, where buyers often have no idea of the value of what they buy, and where middlemen routinely claim reimbursement for sales of things they've never seen to buyers they've never dealt with.* Although the art market connotes mostly Christie's and Sotheby's duopoly, a vast amount of primary and secondary market dealers is present in private galleries, auction houses or on the Internet. Bidding results are often presented by auction houses, whereas prices in private galleries usually remain unpublished. Usage of the Internet, on the contrary, is still relatively small by this conservative market. This paper is focused on auction houses and regards mainly paintings among other collectibles. Here, sold lots are often classified at auctions (like old masters or contemporary art).

Treating art primarily as an asset may be controversial in, generally, two ways. First of all, considering it as an investment may raise some philosophical questions about the purpose of art and monetising *invaluable* features or the relation between an artistic value and a final price. Secondly, studies on the art market have fostered a debate on potential returns. While people may be overwhelmed by an amount of money spent during Christie's or Sotheby's auctions, some researches argue that investing in art yields generally low returns (Baumol, 1986; Frey & Pommerehne, 1989). More recent findings say that returns from art were close to the total return measured by S&P500 between 2003 and 2013. Especially post-war or contemporary and traditional Chinese artworks were among the best investments, delivering compound annual returns of respectively 10.5% and 14.9% (Deloitte & ArtTactic, 2015).

As a consequence of this ongoing debate, there is an increasing number of research initiatives focused on the art market. Different approaches and methods to measure trends, performance and value have been studied. Despite an appraisal based on art history knowledge, a wide range of methods stemming from econometrics and statistics were described in the literature. Data processing, analysis and visualisation have gained special attention in the 21st century and these techniques are applied in a wide variety of domains. Even though relatively unexplored in that manner, the art market is no exception to that trend. Among various machine learning methods, there is one especially used in art market research. Regression analysis is a key technique due to its impact on building price indices describing the whole market. Nevertheless, some other methods are also used.

Identifying trends or measuring the market seems to play the most important role in art market analysis. Art market research on the Internet can follow two approaches. The first considers manually browsing historical sales records, provided by various services. The second, which in fact has its roots in the first approach, relies on various art market indices. In this work we present the current state of art research on methods for description of the art market, taking into account recent developments in econometrics and computer science. Possible further research directions are also outlined.

This paper is organised as follows. First of all, the art market is briefly depicted. Then, Internet auction databases, which can help to minimise the information asymmetry to improve market transparency, are described. The next section sheds light on price indices (commercial as well as those sprung from scientific research) and all underlying calculations. Then, a direct usage of computer science methods for art market analysis is described. The last part of this paper consists of a concise discussion and summary.

Art Price Databases

One of the most straight-forward approaches to invest in art is to pick a potentially interesting artist and check their prices on past auctions. Intuitively, soaring trends may be considered an indicator of a good investment. Although sticking only to increases may not be the best strategy[1], these databases are a valuable source of information. The introduction of Artnet in 1995 increased the number of possible art investors by making auction sales data available (Coslor & Spaenjers, 2013). This section describes some of the most popular art market databases.

A wide range of prices may seem confusing, therefore it is necessary to distinguish each one by introducing the following glossary. *Asking price* is a threshold from which bidding for a given lot starts. Some auction houses provide *estimations* (low and high) to make potential

1 The „Damien Hirst Bubble" may be a good example: http://www.bloomberg.com/bw/articles/2012-11-21/for-collectors-with-hirst-comes-pain.

buyers aware of possible prices. *Reserve price* represents a confidential price agreed upon by the seller and auction house. To sell a lot, at least a set reserve price has to be reached. *Hammer price* is a winning bid price. *Premium price* is equal to the sum of the hammer price and all fees and charges defined by the auction house. During an auction, a lot can be *sold* or *bought in,* which means that there were no bids or it has failed to reach a reserve price.

The following list presents the most popular art market databases:

1) **ArtPrice** Probably the largest database of past auctions, ArtPrice (http://www.artprice.com/) offers access to 30,000,000 prices and indices for about 600,000 artists. Data is gathered from around 4500 auction houses. Hammer prices, estimates, size and auction house information are available only via paid subscription.

2) **Invaluable** This service lists 2,000,000 sold lots and biographic information about 500,000 artists. The data is gathered from approx. the past 15 years (http://www.invaluable.co.uk). All functions are available through paid subscription.

3) **Artnet** Not only does this site play an important role for those who want to buy or sell art, it also contains 9,000,000 auction records (up to 1985), (http://www.artnet.com). Services are offered through paid subscription.

4) **AskArt** This site provides "millions" of auctions results since 1987 and data on over 300,000 artists (http://www.askart.com) via paid subscription.

5) **FindArtInfo** Another site available through paid subscription, FindArt Info provides an access to information on 428,958 artists, 3,662,572 art prices, 349,393 signatures, 2,192,044 photos of artwork and 300 000 signatures (http://www.findartinfo.com).

6) **Blouin Art Sales Index** This site offers around 5,000,000 auction records (http://artsalesindex.artinfo.com). It is worth mentioning that the database is free to browse.

7) **Auction houses pages** It is possible to perform a direct search on numerous auction houses pages. For example, Christie's publishes past auctions results up to 1998. Prior knowledge of a seller during a lot search is a shortcoming of this approach, however.

Additionally, one may can consider following ArtTactic research, a UK-based company that provides financial analysis and profiled reports for art trends for different markets, artists and sectors (http://artsalesindex.artinfo.com).

Indices

The main purposes of calculating price indices are measuring financial performance, evaluating diversification of a potential portfolio and describing trends on the market. Ginsburgh, Mei, and Moses (2006) outline the most important uses for art market indices, which are describing market trends, measuring volatility, and searching for economic incentives influencing market and an appraisal.

Conventional price index methodologies are generally not applicable to non-homogeneous goods (Triplett, 2004). We can distinguish different methodologies for art market index calculation (see Table 1). The simplest and the most popular classification divides them into Repeat-sales Regression (hereafter RSR) and Hedonic Regression (hereafter HR) (Ginsburgh et al., 2006). One may consider different techniques, such as a naïve or composite price index. Naïve indices are based on average and median lot prices in compared years. In this approach it is assumed that distribution of a painting's quality features is constant, which makes it biased. The composite (basket) price index relies on a selection of a representative basket of similar lots and periodic re-evaluation conducted by experts with a domain knowledge (Renneboog & van Houte, 1999). RSR-based indices consider paintings which were sold at least two times to ensure that the same objects are compared. However, this method is criticised due to the relatively small availability of data (only 10% of sold lots (artnet Analytics, 2014)) and long periods between the first and the second sale. On the contrary, hedonic indices can include all traded objects, but they are prone to a selection bias. Notwithstanding, in longer periods and with a richer dataset these indices shall be highly and positively correlated (Ginsburgh et al., 2006).

It is vital to introduce some core econometric concepts behind the art market analysis. "Hedonic" is a word derived from the Greek *hedonikos*, which literally means pleasurable. In terms of economic analysis, these words may be considered as a utility or satisfaction related to consuming given goods (Chau & Chin, 2002). Intuitively, the hedonic regression (HR) relies on "stripping" given goods' characteristics from their price – which, in fact, becomes a function of these characteristics. HR may be used as a method for constructing constant quality price

Table 1: Comparison of indices

Type	Description	Comment
Naive	Naive indices are based on average or median prices. They are good for illustration purposes, but aren't employed in any professional art market analysis.	With an assumption that some paintings' features are constant in time, naive indices are prone to bias.
Composite	Composite indices are based on continuous re-evaluation of some pre-selected lots. Artnet's indices (equations 13 and 14) might be considered partially composite.	This type of index needs employment of experts with domain knowledge. It also might be difficult to calculate them on larger datasets.
Hedonic Regression	Based on a painting's features. It is one of the most popular methods to calculate an index. One can distinguish direct (equation 5) and indirect (equation 9) approaches to building such an index.	Taking into account all sold lots is the main advantage of using hedonic regression in the art market analysis.
Repeat-sales Regression	An index which is built on top of lots sold at least two times. Used in the famous Mei & Moses Fine Art Index (equation 10).	With a sufficient number of observations it can be very accurate, because it compares exactly the same lots. On the other hand, obtaining such data may be very difficult. By its definition, it is also inapplicable to works of newer artists.

Source: Own research

indices, where the considered goods are not traded frequently or data for repeated sales analysis is not available.

Coefficients obtained in the regression analysis can be verified based on e.g. statistical significance, the coefficient of determination or degrees of freedom. Statistical significance means that the p-value is less than a certain level (traditionally 0.05 or 0.01). Generally, the *p-value* is the probability of obtaining a result which is equal to or more extreme than empirical observations, taking into account defined hypotheses. R^2 (coefficient of determination) indicates how well data fits a model obtained in regression analysis. It ranges from 0 to 1, where 1 means a perfect fit. However, very high R^2 is not always desired due to the problem of overfitting (Babyak, 2004).

Although in this paper special attention is paid to art markets, hedonic methods are mostly recognised from their usage in real estate property pricing (Hill, 2013; Hill & Scholz, 2013; Liu, 2014). The extent to which qualities like size, appearance, facilities and features or condition influence the real estate price may be estimated using hedonic methods. However, there are some other non-standard use cases where this approach yields satisfactory results. For example, Moulton (2001) described the expanding role of hedonic methods in the official statistics of the United States of America. Satimanon and Weatherspoon (2010) submitted research on sustainable food products. Kim and Reinsdorf (2013) studied import prices and Wasshausen and Moulton (2006) analysed real Gross Domestic Product. Hedonic regression was even used in retail egg or clothes dryers (http://www.bls.gov/cpi/cpidryer.htm) price analysis.

Early work of Schneider and Pommerehne (1983) on contemporary fine arts describes some assumptions regarding the market. For example, the art market is competitive only to some degree. Moreover, an art gallery owner maximises his (and therefore artists') profits. The authors also proposed equations for supply, demand and existence of an equilibrium on the art market stemming from a hypothesis which says that art price is neither random, nor is it influenced by intangible factors (it may be explained by some supply and demand factors in a statistical way). These assumptions and equations are beyond the scope of this paper, but it is definitely worth mentioning that they described a usage of regression for measuring art markets. Their equation for separating a painting's (i) characteristics from their price P_{it} in year t is presented below:

$$P_{it} = \alpha_0 + \sum_{j=1}^{z} \alpha_j X_{ij} + \varepsilon_{it} \tag{1}$$

which is a standard linear regression, estimated by Ordinal Least Squares (OLS). A painting's characteristics are employed in the equation (1) for X_{ij} as explanatory variables used in a categorical or continuous manner. α_j represents estimated coefficients. The most common characteristics provided by auction houses and used in scientific research as explanatory variables (Kräussl & van Eisland, 2008; Lucińska, 2014) are (for each lot) e.g.: the artist's name, nationality, year of birth, year of death, title of work, year of creation of work, support, technique, height and width, auction house, date of auction, lot number, low and high prior estimate of auction price, signature, sale price, currency of sale price, school, place of sale, works sold in calendar year, average price, publication, number of exhibitions, working periods, or provenance.

Agnello and Pierce (1996) proposed a formula for repeated-sales assets whose value may grow exponentially over time:

$$P_t = Aexp(rt + BX + u_t), t = 0, \dots, T \qquad (2)$$

where t stands for a time period, P_t is a price of a given asset in that period, X is a vector of painting characteristics, r stands for an average rate of return and u_t represents a standard error term. Taking the natural logarithm of both sides we obtain:

$$lnP_t = \alpha + rt + BX + u_t, t = 0, \dots, T \qquad (3)$$

where α = lnA, α+BX is the initial price given a vector of characteristics. Often a limited number of P_t in the available dataset is a disadvantage of this approach. Intuitively, the same lot doesn't get sold in a few years in a row.

Therefore, it was needed to develop a solution which can rely on much shorter periods. A standard approach to show a relation between a given painting and its characteristics considering single hedonic OLS regression is calculated as follows:

$$lnP_{it} = \alpha + \sum_{j=1}^{z} \beta_j X_{ij} + \sum_{t=0}^{\tau} \gamma_t D_{it} + \varepsilon_{it} \qquad (4)$$

where lnP_{it} stands for the natural logarithm of a price of a given painting $i \in \{1,2,\dots,N\}$ at time $t \in \{1,2,\dots,\tau\}$; α, β and γ are regression coefficients for estimated characteristics included in the model. X_{ij} represents hedonic variables included in the model, whereas D_{it} stands for time dummy variables – it is equal to one only if a given painting i was sold in a period t (otherwise it is equal to zero).

Czujack (1997) examined the prices of Picasso's paintings using this approach. Among the most common painting characteristics, the model consisted of additional dummy variables representing so-called "boom periods"(which represent periods of an increased activity on art auction market, as example between 1984 and 1990), provenance, price index, subject and condition. It turns out that neither provenance nor signature plays a statistically significant role in the overall price calculation, as well as pre-auction estimates.

The sculpture market research was performed using the same method by Locatelli-Biey and Zanola (2002). Obviously there was no possibility to apply identical dummy and continuous variables – because of different media or dimensions, for example.

Price indices for the art market can be estimated using a direct or indirect approach in order to track price movements and returns on the art market. A direct approach considers the following calculation:

$$Index_{t+1} = \frac{e^{\gamma_t}}{e^{\gamma_{t+1}}} \qquad (5)$$

where t stands for a considered time period and γ is a regression coefficient obtained in equation (4).

Kräussl and Eisland used the following characteristics of German paintings in their hedonic model (equation (4), X_{ij} variables): surface, type of work, reputation, attribution, living status, and auction house. They proposed a *2-stage hedonic approach* (Kräussl & van Eisland, 2008), an indirect way to calculate price indices in order to eliminate *artist* dummy variables from this model. Regarding β_j coefficients from the equation (4), price indices are calculated as follows:

$$Index = \frac{\prod_{i=1}^{n}(P_{i,t+1})^{1/n} / \prod_{i=1}^{m}(P_{i,t})^{1/m}}{HQA} \qquad (6)$$

Index represents an unweighted geometric mean of painting prices (due to a usually unequal number of paintings in given periods – n and m variables). *HQA* in equation (6) is an abbreviation for *Hedonic Quality Adjustment*. It is calculated to obtain the mean change of paintings characteristics' influence on a price.

$$HQA = exp\left[\sum_{j=1}^{z} \beta_j \left(\sum_{i=1}^{n} \frac{X_{ij,t+1}}{n} - \sum_{i=1}^{m} \frac{X_{ij,t}}{m}\right)\right] \qquad (7)$$

Kräussl and Eisland suggested measuring artistic value with the *True Art Market Index* in a *2-stage hedonic approach*. Putting equations (6) and (7) together results in:

$$Index = \frac{\prod_{i=1}^{n}(P_{i,t+1})^{1/n} / \prod_{i=1}^{m}(P_{i,t})^{1/m}}{exp\left[\sum_{j=1}^{z}\beta_j\left(\sum_{i=1}^{n}\frac{X_{ij,t+1}}{n} - \sum_{i=1}^{m}\frac{X_{ij,t}}{m}\right)\right]} \quad (8)$$

Some changes are brought about in the equation (8) to measure the relative quality corrected value of artist *y*. These changes consider replacing average prices per period $P_{i,t}$ by average prices per artist $P_{i,y}$ and removing artist variables from X_{ij}:

$$Index_y = \frac{\prod_{i=1}^{n}(P_{i,y})^{1/n} / \prod_{i=1}^{m}(P_{i,y-1})^{1/m}}{exp\left[\sum_{j=1}^{z}\beta_j\left(\sum_{i=1}^{n}\frac{X_{ij,t+1}}{n} - \sum_{i=1}^{m}\frac{X_{ij,t}}{m}\right)\right]} \quad (9)$$

As a result, the yielded index can represent an artistic value of a given artist. It resembles an average price per artist's artwork and can replace an *artist* dummy variable in the equation (4). An advantage of this technique is to engender lesser bias which was a consequence of applying *artist* variables to the equation (4).

Kompa and Witkowska (2014; 2015) used a *2-step hedonic regression* presented by Kräussl and Eisland to measure the Polish art auction market. They constructed different models with a relatively high coefficient of determination considering various characteristics. It turns out that *artist alive*, *size* and *price class* are the most important hedonic variables for the final price. The performed hedonic correction was proved to play an important role in indices estimation. Etro and Stepanova (2015) also employed the hedonic index and compared it to the repeated sales price index for the Paris art market between Rococo and Romanticism. The so-called "Death Effect" on painting prices was statistically proven in this paper. Other researchers conducted complex studies on the art indices building. Collins, Scorcu, and Zanola (2009) address the problem of selection bias and time instability in the index by the Heckman 2-stage procedure. Jones and Zanola (2011) argue that usage of the logarithm in HR yields indices that are hard to interpret in an economic way. Bocart and Hafner (2012) suggested a heteroskedastic HR model with a non-parametric local likelihood estimator.

Indices are not only employed in research papers. A wide use of these methods can be observed on commercial websites. To name and describe a few, the following list is provided.

Mei Moses Mei Moses Fine Art Index (http://www.artasanasset.com), founded in 2001, relies on RSR calculated from 30,000 lot records which were sold more than once. The most interesting concerns the fact that the average period between first and second sale is 22 years. However, the index computation relies only on lots sold in Christie's and Sotheby's. Mei and Moses (2002) described this method as follows:

$$r_{i,t} = \mu_t + \eta_{i,t} \quad (10)$$

where *r* stands for a return for a lot *i* in a period *t*, μ_t represents an average return in a period *t*. Therefore, for a lot *i* sold in periods *s* and *b*, the following equation may be provided:

$$r_i = ln\left(\frac{P_{i,s}}{P_{i,b}}\right) = \sum_{t=b_i+1}^{s_i} r_{i,t} = \sum_{t=b_i+1}^{s_i}\mu_t + \sum_{t=b_i+1}^{s_i}\eta_{i,t} \quad (11)$$

The estimated index is calculated by:

$$\hat{\mu} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}r \quad (12)$$

where *X* is a matrix containing rows of dummy variables for each considered lot and Ω stands for a weighting matrix (based on times between sales).

Artnet Despite being a large database, still only 10% of sold lots can be indexed using RSR. This price index is based on a combination of RSR and HR, since RSR can be presented as a nested case of HR (artnet Analytics, 2014). The linear model used is enriched by including *Comparable Sets*, which are in fact the grouped single artist's sales data. These sets are populated taking into account *"appraisal principles and art historical knowledge"*. There are two types of artnet indices: cap-weighted and equal-weighted.

The procedure to calculate the equal-weighted artnet index is as follows:

$$ln(P_{i,s,t}) = \alpha + \frac{1}{N_S}\sum_{t=1}^{T}\sum_{j=1}^{N_{st}} ln(P_{j,s,t}) + \sum_{t=1}^{T}\gamma_t D_{it} + \sum_{s=1}^{S}\delta_s C_{is} + \varepsilon_{ist} \quad (13)$$

where $P_{i,s,t}$ stands for a price of a lot *i (i = 1,...,N)* in a Comparable Set *s* in a time *t (t = 1,...,T)*. A Comparable Set is denoted as $C_{i,s}$ *(s = 1,...,S)* and N_s stands for the total amount of lots belonging to $C_{i,s}$ A corrected price is defined as $Y_{i,s,t}$:

$$Y_{i,s,t} = ln(P_{i,s,t}) - \frac{1}{N_S}\sum_{t=1}^{T}\sum_{j=1}^{N_{st}} ln(P_{j,s,t}) = \alpha + \sum_{t=1}^{T}\gamma_t D_{it} + \sum_{s=1}^{S}\delta_s C_{is} + \varepsilon_{ist} \quad (14)$$

The parameters of the model are estimated by OLS:

$$E(Y_{i,s,t}) = E\left(\alpha + \sum_{t=1}^{T} \gamma_t D_{it} + \sum_{s=1}^{S} \delta_s C_{is}\right) \quad (15a)$$

$$E(Y) = E(XB) \quad (15b)$$

$$B = (X'X)^{-1}(X'Y) \quad (15c)$$

The index (with the base value set to 100) is calculated as follows:

$$Index_t = 100e^{\gamma_t} \quad (16)$$

The cap-weighted artnet index puts emphasis on high valued lots. The procedure is roughly the same as in the equal-weighted index, however the equation (17c) is used instead of (15c):

$$\Omega_{i,s,t} = \frac{\frac{1}{N_{s,t}} \sum_{j=1}^{N} P_{j,s,t}}{\sum_{s=1}^{S_t} \sum_{j=1}^{N_{s,t}} P_{j,s,t}} \quad (17a)$$

$$W = diag(\Omega_{1,1,1}, \dots, \Omega_{N,S,T}) \quad (17b)$$

$$B = (X'WX)^{-1}(X'WY) \quad (17c)$$

Note the similarity (a usage of the weighting matrix) to the equation (12) used in Mei Moses Fine Art Index.

Artprice Not only does Artprice provide the price database, it also presents its index for different artists. However, it is not precisely described: *"the first step is to quantify the effect of quality on price. This analysis is not based on the work itself but on the various elements that go into it. Works of art are defined by a matrix of features. Using these estimates we deduce a set of equations that quantify the influence of each characteristic. From these we can create synthetic works for each artist. Having estimated the parameters it is possible to find the average price of a composite work for each period. All auctions relating to an artist are included in the calculation thus avoiding any mistakes from selectivity. The data universe is fully inclusive"* (source: © Artprice.com http://www.artprice.com/).

Other examples are, for instance, the Art Market Research Index (http://www.artmarketresearch.com) (available via paid subscription) or the Citadel Art Price Index (https://www.citadel.co.za/ArtIndex/469/art-price-index). The first one is worth mentioning due to the possibility of creating own indices based on the selected parameters (like a market segment or inflation).

Although hedonic and repeat-sale methods are probably the most popular approaches used in art market analysis, it is worth mentioning that some other techniques were described in the literature. For example, Førsund and Zanola (2006) employed the *Data Envelopment Framework (DEA)* model to measure the impact of auction houses on the hammer price for Picasso paintings. The DEA technique comes from operations research. Additionally, Charlin and Cifuentes (2014) introduced a new financial metric for the art market – *Artistic Power Value (APV)*. It is based on the price per unit of area of painting.

Employment of IT in Art Market Analysis

The art market is a challenging domain to provide analytical IT support. On one hand, it is a data rich domain with a number of data sources available. On the other, this data is distributed, following different data models, often not publicly available (or available to a certain extent). Moreover, analyses cannot follow typical approaches, because of the specifics of the market that were discussed earlier in this paper.

IT support can be therefore described from two points of view: how to deal with the data describing the art market as well as what analyses can (or should) be carried out. This section will try to provide answers to these questions.

The typical approach to data processing towards its analysis follows the process:

1) data collection,
2) data integration and enrichment (if any),
3) data analysis,
4) visualisation of results.

In case of the art market we can apply the same procedure, however it is worth discussing challenges that occur.

The data on the art market that may support further analyses is available in many sources. These sources cover e.g.: art price databases (discussed before); data on artists on various portals, Wikipedia, blogs, social media, personal pages, etc.; data on market trends; results of auctions; various experts' analyses. This catalogue is not by any means complete. The major challenges here concern the data diversity and its volatility that influence techniques of crawling, filtering and retrieval.

The second step of data processing relates to the data integration and cleansing. Taking into account various data models, formats, completeness, comparability, influenced by techniques applied in the previous step, the topic of data integration is a huge challenge that may be addressed using e.g. the Linked Open Data standards. This phase may be further extended while applying data enrichment methods i.e. to expand the data models.

The two previous phases are aimed at the preparation of data for analyses. In the following step, numerous techniques that enable making value out of data might be applied e.g. statistical, econometrical, data mining, etc. The techniques applied depend on the goal of the analysis to be carried out. Some of them were discussed in the paper (IT support for econometrics, data mining or machine learning) and are the subject of the ongoing research.

The last step concerns the visualisation that delivers outcomes of analyses held or enables browsing of raw data retrieved. Usually, this concerns presentation charts, clusters, trends, relations (including graphs), etc. that may enable drawing conclusions regarding the subject of analysis.

On the contrary to the data processing procedure, which is uniform for many domains, the types of analyses to be held are mostly specific for the art market.

The first type of research on the art market concerns fraud. Collecting and analysing data that may prove authenticity of art works (signatures, techniques, quality, etc.), origin, fraudulent transactions and even might enable prevention of fraud on the market are of importance. This is usually the issue of the smaller markets or transactions that are not supported by professional auction houses.

The second area that should be the subject of analyses concerns inflated prices and false market trends. The current inability to provide data that may validate the prices on the art market or support individuals in the "decision making process, strengthens the role or "experts" being in fact creators of an "investment bubble". This relates also to the issue of reputation addressed by various authors.

Canals-Cerda (2008) analysed the value of a good reputation at selling artwork online. The previously conducted research shows that there is a small impact of a seller's reputation on the sale price, but it has a negative influence on the number of bidders. On the contrary, reputation in selling art plays an important role according to this study. An empirical analysis was conducted on data provided by eBay, a leading on-line auction service.

Renneboog and Spaenjers (2013) in their HR model employed information from books and Internet data sources (e.g. "Gardner's Art through Ages") in order to describe the artist's reputation. According to this article: the more occurrences of an artist's name in the corpus, the better his reputation is.

Another field where application of various data processing methods may bring value concerns comparison of indices between markets. The challenge concerns the number of indices, diverse methods to quantify these indices (and their incomparability) as well as the number of domains and of markets (even in the globalisation era).

An interesting direction of research is also the application of Linked Open Data to enrich the data available on artists, their art works and the market to provide more high quality data subject for further study. Storing all data using Linked Data Standards may also help to address the data integration challenges mentioned previously.

Analysis of popularity based on social media i.e. Facebook, Instagram or Twitter also is gaining increasing attention. The rising interest (reflected in a number of mentions on social media) concerning intangible goods, usually relates to increasing prices of these goods. Monitoring trends on social media may therefore influence the investment decisions.

Coslor and Spaenjers (2013) in their paper e.g. used Google's corpus of books. This raw data is publicly available (http://storage.googleapis.com/books/ngrams/books/datasetsv2.html). The aim of this research was to find a correlation between the interest in art investment and the use of the art investment lexicon. The applied methodology considers the frequency analysis of so-called n-grams (sequences of n words) over the Google Books corpus. Obtained results had not been self-explanatory. Therefore, it was necessary to perform a qualitative analysis to obtain better conclusions. Although that paper was focused on the growth of the art investment field, it shows a usage of a simple text analysis over large datasets to better understand the subject matter.

Elgamal and Saleh (2015) came up with a computational framework as well as carried out research on artworks' creativity (which is based on originality and influence). They also conducted a "time machine" experiment, which was used to measure how creativity of artworks has changed over centuries.

Last but not least, IT is also employed in the area of digitalisation of art that is out of the scope of our research, but could provide an interesting input to image processing and identification of fraud described briefly in this paper.

To summarise, numerous services and auction houses are publishing art market historical data, others are even suggesting which pieces of art should be invested in. Currently, there is one service which offers a completely different approach to art investment. ArtRank (http://artrank.com) (formerly known as SellYouLater), a controversial (http://www.theguardian.com/artanddesign/2014/jun/23/artrank-buy-sell-liquidate-art-market-website-artists-commodities), (http://www.nytimes.com/2015/02/08/magazine/art-for-moneys-sake.html?_r=0) start-up, tries to highlight potentially profitable artists on a quarterly basis by using machine learning algorithms (Velthuis, 2014). Due to its commercial nature, the mechanisms behind this project are not described in detail. A brief explanation on the ArtRank's page says: *"The algorithm is comprised of six exogenous components: Presence, Auction results, market Saturation, market Support, Representation and Social mapping (PASSRS). Each component is qualitatively weighted in service of defining a vector or 'artist trajectory'. We compare past trajectories to help forecast early emerging artists' future value. (...) Our purpose-coded machine-learning algorithm extracts relevant explanatory metrics from over three million historic data points including auction results, representation, collectors, and museums. These weighted qualitative metrics work in conjunction with our classification algorithm to identify prime artist prospects based on known trajectory profiles"*. Although PASSRS and artist trajectories are not explained in any way, this site definitely grabs attention and may be a source of further investigation.

Discussion and Summary

This paper presented a short introduction to art market analysis regarding techniques stemming from econometrics and computer science. Topics such as Internet data sources, indices construction and employment of information technology on art market research were covered. Areas of possible future exploration are outlined in the remainder of this section.

After this analysis, a couple of future research questions and problems can be addressed. First of all, a further development and evaluation of different indices used in art markets is needed. Although various methods have been presented so far and a foundation is given, to the best of our knowledge there is no agreement on an ultimate and robust method to measure the art market. The presented equations can yield incomparable results. Selection bias and explanatory variables issues were addressed in some research. While the focus is mainly on statistical and econometric areas, only a few research papers attempt to enrich the set of "standard" explanatory variables. Future research may include an adoption of concepts related to ontologies or the Semantic Web, which should bring about ample, structured and extensible sets of characteristics.

Due to an increasing usage of social media platforms among art collectors (ArtTactic, 2015), it also seems reasonable to include these sources of data in the art market research. Some of the presented methods consider evaluation conducted by experts with domain knowledge (art historians in this case). Machine learning techniques could be an inspiration for future research, especially taking into account ArtRank philosophy, en route to better understand processes which are shaping the market. Another possible example is creating comparable sets in artnet price indices automatically.

References

Agnello, R.J., Pierce, R.K. (1996). Financial Returns, Price Determinants, and Genre Effects in American Art Investment. *Journal of Cultural Economics*, 20(4), 359–383.

artnet Analytics, (2014). *artnet Indices White Paper* (Tech. Rep.).

ArtTactic, (2015). *The Hiscox Online Art Trade Report 2015* (Tech. Rep.).

Babyak, M. (2004). What You See May not be What You Get: a Brief, Nontechnical Introduction to Overfitting in Regression-type Models. *Psychosomatic medicine*, 66(3), 411-21.

Baumol, W. (1986). Unnatural Value: Or Art Investment as Floating Crap Game. *American Economic Review*, 76(2), 10–14.

Bocart, F.Y., Hafner, C.M. (2012, November). Econometric Analysis of Volatile Art Markets. *Computational Statistics & Data Analysis*, 56(11), 3091–3104.

Canals-Cerda, J.J. (2008, April). The Value of a Good Reputation Online: An Application to Art Auctions. SSRN *Electronic Journal*.

Charlin, V., Cifuentes, A. (2014). A New Financial Metric for the Art Market. Munich.

Chau, K., Chin, T.L. (2002, June). A Critical Review of Literature on the Hedonic Price Model. *International Journal for Housing Science and Its Applications* (27), 145–165.

Collins, A., Scorcu, A., Zanola, R. (2009, August). Reconsidering Hedonic Art Price Indexes. *Economics Letters*, 104(2), 57–60.

Coslor, E., Spaenjers, C. (2013, November). Organizational and Epistemic Change: The Growth of the Art Investment Industry. *HEC Paris Research Paper*, FIN-2013-1018.

Czujack, C. (1997). Picasso Paintings at Auction, 1963–1994. *Journal of Cultural Economics,* 21(3), 229–247.

Deloitte & ArtTactic, (2015). *Art & Finance Report 2014* (Tech. Rep.).

Elgamal, A., Saleh, B. (2015). Quantifying Creativity in Art Networks. *International Conference on Computational Creativity (ICCC).*

Etro, F., Stepanova, E. (2015). *The Market for Paintings in Paris between Rococò and Romanticism. Kyklos* 68(1), Venice, 68-1.

Førsund, F.R., Zanola, R. (2006). DEA Meets Picasso: The Impact of Auction Houses on the Hammer Price. *Annals of Operations Research,* 145(1), 149–165.

Frey, B.S., Pommerehne, W.W. (1989). Art Investment: An Empirical Inquiry. *Southern Economic Journal,* 56(2), 396–409.

Ginsburgh, V., Mei, J., Moses, M. (2006). The Computation of Prices Indices. In *Handbook of the economics of art and culture* (Vol. 1, pp. 947–979).

Hill, R.J. (2013). Hedonic Price Indexes for Residential Housing: A Survey, Evaluation and Taxonomy. *Journal of Economic Surveys*, 27(5), 879–914.

Hill, R.J., Scholz, M. (2013). Incorporating Geospatial Data into House Price Indexes: A Hedonic Imputation Approach with Splines. *Graz Economic Papers*, 2014-05.

Jones, A.M., Zanola, R. (2011). Retransformation bias in the Adjacent Art Price Index. *ACEIWorking Paper Series*, (1).

Kim, M., Reinsdorf, M. (2013). *A Test of Hedonic Price Indexes for Imports.*

Kompa, K., Witkowska, D. (2014). Indeks Hedoniczny Malarstwa Polskiego dla Najbardziej Popularnych Autorów na rynku aukcyjnym w latach 2007-2010. *Acta Universitatis Nicolai Copernici*, 1(1), 7–26.

Kräussl, R., van Eisland, N. (2008). Constructing the True Art Market Index - A Novel 2-Step Hedonic Approach and its Application to the German Art Market. *CFS Working Paper Series*, 2008/11.

Liu, X. (2014). Comparison of Hedonic and Repeat-Sales House Price Indexes: Turning Points, Appreciating Rates and Sample Bias.

Locatelli-Biey, M., Zanola, R. (2002). The Sculpture Market: An Adjacent Year Regression Index. *Journal of Cultural Economics,* 65–78.

Lucińska, A. (2014, December). The Art Market in the European Union. *International Advances in Economic Research*, 21(1), 67–79.

Mei, J., Moses, M. (2002). Art as an Investment and the Underperformance of Masterpieces. *American Economic Review*, 92(5), 1956-1668.

Moulton, B.R. (2001). The Expanding Role of Hedonic Methods in the Official Statistics of the United States. *BEA Papers*, 0018.

Plattner, S. (1998). A Most Ingenious Paradox: The Market for Contemporary Fine Art. *American anthropologist*, 100(2), 482–493.

Renneboog, L., Spaenjers, C. (2013). Buying Beauty: On prices and Returns in the Art Market. *Management Science*, 59(1), 36-53.

Renneboog, L., van Houte, T. (2002). The Monetary Appreciation of Paintings: From Realism to Magritte. *Cambridge Journal of Economics*, 26, 331-57.

Satimanon, T., Weatherspoon, D.D. (2010). Hedonic Analysis of Sustainable Food Products. *International Food and Agribusiness Management Review*, 13(4), 57–74.

Schneider, F., Pommerehne, W.W. (1983). Analyzing the Market of Works of Contemporary Fine Arts: An Exploratory Study. *Journal of Cultural Economics*, 7(2), 41–67.

Triplett, J. (2004). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application To Information Technology Products*. Paris.

Velthuis, O. (2014). ArtRank and the Flippers: Apocalypse Now? *Texte zur Kunst*, 96, 34-49.

Wasshausen, D., Moulton, B.R. (2006). The Role of Hedonic Methods in Measuring Real GDP in the United States. *BEA Papers*, 0067.

Witkowska, D., Kompa, K. (2015). Constructing Hedonic Art Price Indexes for the Polish Painting Market using Direct and Indirect Approaches. *AESTIMO, The IEB International Journal of Finance, 10,* 2–25.