T E C H N I C A L
R E P O R T

11031

Econometric analysis of volatile art markets

BOCART, F. and C. M. HAFNER

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Econometric analysis of volatile art markets

Fabian Y. R. P. Bocart* Christian M. Hafner†

June 2011

Abstract

A new heteroskedastic hedonic regression model is suggested. It takes into account time-varying volatility and is applied to a blue chips art market. Furthermore, a nonparametric local likelihood estimator is used. This estimator is more precise than the often used dummy variables method. The empirical analysis reveals that errors are considerably non-Gaussian, and that a student distribution with time-varying scale and degrees of freedom does well in explaining deviations of prices from their expectation. The art price index is a smooth function of time and has a variability that is comparable to the volatility of stock indices.

Keywords: Volatility, art markets, hedonic regression, semiparametric estimation

*Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, Belgium.

†CORE and Institut de statistique, biostatistique et sciences actuarielles, Université catholique de Louvain, 20, voie du roman pays, 1348 Louvain-La-Neuve, Belgium; telephone: 0032 10 47 43 06; email: `christian.hafner@uclouvain.be`, corresponding author.

1 Introduction

It is well documented that volatility of many commodities and stocks displays a certain degree of time variation. This feature has important consequences for economists, policy makers, economic agents, actors in the financial and commodity markets. Since Engle (1982)'s ARCH model, several models have been built to investigate volatility of commodities, and a large literature is now dedicated to its time-varying structure.

Surprisingly, while considerable efforts have been devoted to assess returns in the art market, few studies attempt to investigate the volatility structure of art as a function of time. Yet, volatility of fine art is worth investigating, and a better understanding of its structure may be of practical use for market participants, more particularly for participants exposed to derivatives on art. Such derivatives include price guarantees underwritten by auction houses (Greenleaf et al., 1993) that are similar to short positions in put options. Volatility of fine art also plays a role when pieces of art are used as collateral for loans (McAndrew and Thompson, 2007). Campbell and Wiehenkamp (2008) illustrate the mechanism of another art-based option: the Art Credit Default Swap: A bank lends money to an entity on the one hand, and buys an option (the Art Credit Default Swap) from a third party -the seller of protection- on the other hand. This option gives the bank the right to swap the art object against cash, would the borrower default. Many other derivatives, sensitive to volatility, abound in the market for physical insurance on luxury goods and art. Unlike commodities exchanged on organised platforms, a common complication in analysing the market for art and antiques is the heterogeneity of exchanged goods. This feature prevents the observer from directly estimating returns and volatility of the market. As far as returns are concerned, two main methodologies have been developed to cope with this issue: the repeat sale methodology (RSM) and the hedonic regression. RSM is based on various goods that have been sold several times in different periods, so as to compute an average rate of return. RSM has been used by Baumol (1986), Goetzmann (1993), Pesando (1993), as well as Mei and Moses (2002). A major critique against RSM is that it focuses on a small, biased sample of goods (Collins et al., 2009) that have been resold through time.

This paper focuses on the hedonic regression methodology (HRM) that is further detailed in Section 2. Hedonic regression has been favoured to study the art market by Chanel et al. (1994), Hodgson and Vorkink (2004), Collins et al. (2009), Oosterlinck (2010), Renneboog and Spaenjers (2009) and Bocart and Oosterlinck (2011). Hedonic regression has the advantage of using all goods put for sale. The approach is to regress a function of the price of each good on its characteristics, including time dummy variables whose coefficients will constitute the basis for building an index. The main disadvantage of hedonic regression methodology is that the index depends on the explanatory variables. Ginsburgh et al. (2006) discuss the main problems of hedonic regression applied to the art markets, such as the choice of a functional form, the specification bias and the "revision volatility" -that is, as new data are included in the dataset, the price index changes. Methodology-wise, ordinary least squares are usually employed to estimate parameters. Recent research aims at correcting methodological flaws in hedonic regression: Collins et al. (2009) introduce the Heckman procedure to take into account a sample selection bias linked to unsold artworks as well as a Fisher index to cope with time instability of parameters. Jones and Zanola (2011) detail the use of a so-called smearing factor to correct for a retransformation bias when a log scale of prices is used in the regression. Scorcu and Zanola (2010) suggest a quantile regression to take into account the fact that parameters depend on price levels. Hodgson and Vorkink (2004), highlight that for the art market, non-Gaussianity is an issue that needs to be treated, since OLS estimates are not efficient. They assume an i.i.d. error term with nonparametric density function, and suggest Bickel's adaptive estimation to obtain efficient estimates. While this is an important improvement of standard OLS estimation in this framework, the assumption of i.i.d. errors may seem too restrictive for markets which exhibit time-varying features such as changing uncertainty concerning the evaluation of art. In particular, we show in this paper that art markets can be heteroskedastic.

We recommend a local maximum likelihood procedure to obtain time-varying estimates of higher moments, i.e., variance, skewness and kurtosis. The time-varying variance is later used to derive what we call "volatility of predictability". It can also be used to obtain more efficient parameter estimates by using weighted least squares. However, our main interest lies in volatility in itself, as it can be used further e.g. for derivative pricing. Modelling unconditional volatility as a deterministic function of time has become popular recently in financial markets, starting from Engle and Rangel (2008) who use a spline estimator for unconditional volatility combined with a classical GARCH model for conditional volatility. Our research follows the same spirit but allows moreover for time-varying skewness and kurtosis.

The paper is organised as follows: Section 2 introduces the data we use to build a blue-chips art index and presents the HRM methodology and a time-dependent estimator for variance. Section 3 illustrates our results. Concluding remarks are presented in section 4.

2 Data and methodology

2.1 Data

We choose to restrict our analysis to two-dimensional artworks, excluding works on paper and photographs, made by artists ranked amongst the top 100 sellers (in sales revenue in auction houses, according to Artprice, a company specialized in publishing auction results), both in 2008 and 2009. The rationale behind this choice is that large volumes of sales may signal a particular interest from the market for these artists.

We believe that these artists are more likely to be seen both as consumption goods and investment goods unlike many little traded artists whose objects are more likely to be bought as pure consumption goods. Indeed, Frey and Eichenberger (1995) state that "pure speculators" who consider art as an investment may avoid markets presenting too much uncertainty (such as financial risk or attribution risk).

Furthermore, Goetzmann (1993) emphasizes that art prices are influenced by "stylistic risk", that is the risk of having not enough bidders when reselling the artwork. Mei and Moses (2002) compare stylistic risk in art markets to liquidity risk in financial markets. Unknown and relatively little traded artists are typically cursed by considerable financial and liquidity risk, as it can be difficult to realize a sale in a market where demand is weak.

On the other hand, buyers of liquid artists – with a low stylistic risk – know ex-ante that they will be able to re-sell artworks, which might attract speculators and investors. In practice, art is actually traded as an investment. This is empirically confirmed by activity from dealers, funds, foundations and private individuals who store artworks in warehouses, bank vaults, or in Switzerland's port-franc containers, where obviously the aesthetic return is null.

Based on the assumption that liquid artists can be seen as an investment, we focus on "Blue Chips Artists": we need to select artists who stay in the top 100 of best sellers two consecutive years, in order to avoid bias from exceptional or unusual sales. Forty artists correspond to this description, out of which 32 stayed in the top 100 from 2005 to 2010 in a row. We record auction data from January 2005 to June 2010. 5612 sold pieces are recorded. An auction process is an opportunity to record information. Auction houses announce weeks to months in advance the dates when auctions will occur. Sometimes, a single auction is split into several days. In most cases, the sale is organized around a certain theme ("Impressionist art" for instance). Prior to the auction, a catalogue is published by the auction house. In this catalogue, each artwork is linked to a lot number, a price estimate, and a description. The length of the description differs from one artwork to another, but key variables are systematically recorded. For each sale, we gather the following information: the price in USD, and whether it is a hammer price (that is, the price reached at auction), or a premium price (the price including the buyer's premium), the sales date, the artist's name, the width and height of the painting in inches, the year it has been made, the painting's title, the auction house and city where the sale occurred and the title of the auction house's sales theme. From this information we extract additional variables, such as the subject of the painting (derived from the title), the theme of the auction (modern, contemporary, impressionist, etc.), derived from the sale's title, the artist's birthday, at what age he painted the piece and whether he was alive or dead at the time of the auction. We also derive the weekday of the sale. Some factors are omitted that may influence the final price for a painting, such as exhibition costs, transaction costs, and transport. All variables are presented in tables 3, 4 and 6.

2.2 Hedonic Regression Methodology

Hedonic regression is a common tool to estimate consumer price indices (see e.g. Ginsburgh et al., 2006) and has been widely used in real estate and art markets. Let p_i denote the price of sale i. The logarithm of this price is usually modelled by the following hedonic regression model,

$$\log p_i = \nu + \sum_{t=1}^{T} \beta_t d_{i,t} + \sum_{k=1}^{K} \alpha_k v_{i,k} + u_i, \quad i = 1, ..., N. \tag{1}$$

$d_{i,t}$ is a dummy variable taking the value 1 if the artwork i was sold in period t, and 0 otherwise. ν is a constant term. The time index $t = 1$ corresponds to the very first period of the series and is used as benchmark. In our case, it would be the first quarter of 2005. For identification, we set β_1 equal to zero. The K variables $v_{i,k}$ are all other characteristics of the piece of art i (for instance: the height, surface, and dummies for the artists, subject, etc.). The index, with base 100 in $t = 1$, using a bias correction factor based on Duan (1983) is then defined as follows, see Jones and Zanola (2011):

$$\text{Index}_t = 100 \times e^{\beta_t} \times \frac{\frac{1}{N_t}\sum_{i=1}^{N} d_{i,t} e^{\hat{u}_i}}{\frac{1}{N_1}\sum_{i=1}^{N} d_{i,1} e^{\hat{u}_i}}, \tag{2}$$

where $N_t = \sum_{i=1}^{N} d_{i,t}$ is the number of observations at time t. Regression (1) is generally estimated using Ordinary Least Squares (OLS). OLS estimators are efficient when errors u_i are normally distributed with constant variance, i.e., $u_i \sim N(0, \sigma_u^2)$. Data from art sales, however, often violate this assumption. Hodgson and Vorkink (2004) and Seckin and Atukeren (2006) focus on the normality part and propose a semiparametric estimator of the index based on a nonparametric error distribution, while maintaining the assumption that u_i is i.i.d. and, hence, homoskedastic.

Furthermore, indices based on the OLS methodology suffer from a sample selection bias. Indeed, only sold paintings are taken into account, whereas unsold paintings carry important information as well. Collins et al. (2009) suggest a two-stage estimation to cope with the issue. Let S_i denote a dummy variable taking value 1 if the painting i has been sold and 0 otherwise. The first stage involves a probit estimation:

$$P(S_j = 1 \mid w_j) = \Phi\left(\sum_{p=1}^{P} \delta_p w_{j,p}\right),\ j = 1, ..., N+U, \tag{3}$$

where Φ is the cumulative distribution of the standard normal, N is the number of pieces sold and U is the number of unsold pieces. The P variables $w_{j,p}$ are characteristics of the piece of art j (for instance: the height, surface, and dummies for the artists, subject, etc.), and $\delta = (\delta_1, \ldots, \delta_P)'$ is a parameter vector.

The second stage involves an OLS estimation similar to equation (1), but only for the sold pieces ($S_i = 1$):

$$\log p_i = \nu + \sum_{t=1}^{T} \beta_t d_{i,t} + \sum_{k=1}^{K} \alpha_k v_{i,k} + \kappa \zeta_i + u_i,\ i = 1, ..., N. \tag{4}$$

The term ζ_i is a correcting variable, based on parameters of the probit estimation and found using the procedure of Heckman (1979).

We now propose to modify the time component, replacing the time dummies $d_{i,t}$ by a smooth unknown function of time, and allowing for heteroskedasticity of unknown form. An important advantage of choosing a continuous function $\beta(t)$ rather than time dummies is that one avoids gathering paintings sold at different periods in a single variable. We also remove the normality assumption, allowing for skewness and leptokurtosis. In particular, we assume that residuals are distributed according to a student-skewed distribution. Our semiparametric heteroskedastic model can then be written as

$$\log p_i = \nu + \sum_{k=1}^{K} \alpha_k v_{i,k} + \kappa\zeta_i + \beta(t_i) + \sigma(t_i)\varepsilon_i, \quad i = 1, ..., N, \tag{5}$$

or, alternatively:

$$\log p_i = \sum_{m=1}^{M=2+K} \gamma_m x_{i,m} + \xi_i, \quad i = 1, ..., N, \tag{6}$$

where $x_i = (1, v_{i,1}, ...v_{i,k}, ..., v_{i,K}, \zeta_i)$, and

$$\xi_i = \beta(t_i) + \sigma(t_i)\epsilon_i = \beta(t_i) + u_i, \quad u_i = \sigma(t_i)\epsilon_i. \tag{7}$$

The function $\sigma(t)$ is a smooth function of time, t_i is the selling time of the i-th sale, $\beta(t)$ is the trend component of the log price at time t, and for identification we restrict its mean to zero. The error term ε is independent, but not identically distributed, with mean zero and variance one, given by a standardized student skewed distribution. The probability density function of the student skewed distribution $t(\eta, \lambda)$ with mean zero and variance equal to one is provided by Hansen (1994):

$$g(\varepsilon \mid \lambda, \eta) = bc\left(1 + \frac{1}{\eta - 2}(\frac{\varepsilon + a}{1 - \lambda})^2\right)^{\frac{-(\eta+1)}{2}} \quad \forall \varepsilon < -a/b, \tag{8}$$

and

$$g(\varepsilon \mid \lambda, \eta) = bc\left(1 + \frac{1}{\eta - 2}(\frac{\varepsilon + a}{1 + \lambda})^2\right)^{\frac{-(\eta+1)}{2}} \quad \forall \varepsilon \geq -a/b, \tag{9}$$

where η stands for the degrees of freedom with $2 < \eta < \infty$, and λ is a parameter characterizing the skewness of the distribution, with $-1 < \lambda < 1$.

The constants are given by

$$a = 4\lambda c\frac{\eta-2}{\eta-1}, \quad b^2 = 1+3\lambda^2-a^2, \quad c = \frac{\Gamma(\frac{\eta+1}{2})}{\sqrt{\pi(\eta-2)}\Gamma(\frac{\eta}{2})}. \tag{10}$$

A first stage estimation of γ is a prerequisite. We suggest constructing feasible weighted least squares (FWLS) estimators of γ:

$$\hat{\gamma} = (X'\hat{W}X)^{-1}X'\hat{W}Y, \tag{11}$$

where X is the N×M matrix of observed independent variables, Y is the $N \times 1$ vector of observed log-prices, $\hat{W}$ is an N×N diagonal matrix with $w_{ii} = \hat{\sigma}^{-2}(t_i)$. If a nonparametric Nadaraya-Watson estimator is used for $\hat{\sigma}^2$, then the estimator (11) has been first proposed by Rose (1978). For the case of a volatility function depending on an i.i.d. random variable, Carroll (1982) showed that it is asymptotically equivalent to the WLS estimator with known volatility function. We can consistently estimate the variance of the FWLS estimator by

$$\widehat{\mathrm{Var}}(\hat{\gamma}) = (X'\hat{W}X)^{-1}. \tag{12}$$

Because it yields more precision in parameter estimates, FWLS may lead to a better selection of explanatory variables, as compared to the OLS methodology. This is important since *"choosing the functional form and the variables that represent quality are pervasive in hedonic indexing, and can lead to all the problems linked to mis-specification"* (Ginsburgh et al., 2006).

Conditional on this first stage estimate of γ, we use a nonparametric estimation, introducing a kernel function K and a bandwidth h. We suggest estimating the local parameter vector $\theta = (\beta, \sigma, \eta, \lambda)'$ by local maximum likelihood. One advantage of considering θ as a function of continuous time is the improved stability of estimation compared to ordinary least squares with time dummies. Indeed, we avoid all risks linked to the inversion of a near singular matrix, a problem often met when few data are available in a given period. Smoothing over several adjacent time periods allows to stabilize the estimation of a parameter at a given time. Formally, the local likelihood estimator of θ is defined as

$$\widehat{\theta}(\tau) = \mathrm{argmax}[l(\theta \mid \xi, \tau, h)], \tag{13}$$

where

$$l(\theta \mid \xi, \tau, h) = \sum_{i=1}^{N} \log[g_s(\xi_i \mid \theta)]K(\frac{t_i-\tau}{h}), \tag{14}$$

$$g_s(\xi_i \mid \theta) = \frac{1}{\sigma} g\left(\frac{\xi_i - \beta}{\sigma} \mid \lambda, \eta\right), \tag{15}$$

and where $g(\cdot)$ is the standardized skewed student t density given in (8) and (9). No closed form solution for (13) is available in the general case, but numerical methods can be employed in a straightforward way to maximize the local likelihood function and thus obtain local parameter estimates.

In order to construct pointwise confidence intervals, we proceed as in Staniswalis (1989). Let υ denote one of the four local parameters $(\beta, \sigma, \eta, \lambda)$ and ι, the three others. An expression for the variance $\mathrm{Var}(\hat{\upsilon})$ is given by

$$\mathrm{Var}(\hat{\upsilon}) = \frac{\| K \|^2}{NhI(\upsilon)f(t)}, \tag{16}$$

where

$$I(\upsilon) = E\left\{\left(\frac{\partial log[g_s(\upsilon \mid \iota)]}{\partial \upsilon}\right)^2 \mid u\right\}, \tag{17}$$

$f(t)$ is the density of the time of sales, and $\| K \|^2$ is the L_2 norm of the kernel used in equation (14). Based on the asymptotic normality of the estimator of $\theta(\tau)$, one can then construct pointwise confidence intervals as usual.

The special case where $\lambda(t) = 0$ and $\eta(t) = \infty$ yields the Gaussian likelihood, for which the maximizer is available in closed form and given by the Nadaraya-Watson estimator (Hardle, 1990). Hence, our estimator nests the Nadaraya-Watson estimator as a special case.

Bandwidth selection can be based on a classical plug-in methodology for bandwidth selection, following Boente et al. (1997):

$$h = N^{-1/5} \frac{\| K \|^2 \sigma^2}{C_2^2(K) \int_0^1 m''(u)^2 du}, \tag{18}$$

where $C_2(K) = \int_{-\infty}^{+\infty} x^2 K(x) dx = 1$, $m''(u) = \sum_{i=1}^{n} \frac{1}{Nh_0^3} K''(\frac{u-u_i}{h_0}) u_i$, σ^2 is the empirical variance of ξ and h_0 is a pilot bandwidth. We follow this procedure in the empirical example of the following section.

3 Results

3.1 Hedonic Regression

We first build a quarterly index using time dummies, using an OLS methodology with Heckman correction. The variables selected in the probit equation

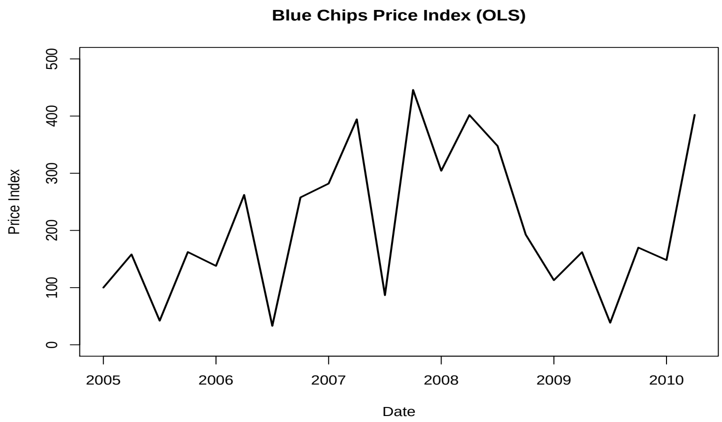

Figure 1: Price index resulting from equation (2), based on time dummies and estimated by Ordinary Least Squares.

.

(3) are presented in Table 2. Variables included in regression (4) are selected following a backward selection methodology: they are kept in the model if significant at a level of 5% using OLS regression. Advantages and disadvantages of backward selection are discussed e.g. in Hendry (2000). As compared to other selection methodologies such as forward selection, backward selection suffers from the fact that the initial model may be inadequate. Indeed, non-orthogonality of variables may lead to erroneously eliminate variables, or wrongly keep colinear variables. To avoid this problem, we run different initial models, separating variables that share a high degree of colinearity. The model presenting the highest adjusted-R^2 has been kept.

Table 1 summarizes results from the regression and Figure 1 presents the resulting OLS-based price index. The need to correct for time dependent error variance is indicated by a Breusch-Pagan test for heteroskedasticity on OLS residuals, which delivers a p-value of 0.02. We hence reject the null hypothesis of homoskedasticity at a level of 5%. The quantile plot in Figure 7 highlights that normality of residuals is an unrealistic assumption.

We then proceed with our methodology: we discard time-dummy variables and select explanatory variables with a backward selection methodology

Table 1: Parameters estimates of regression 5. Variables are selected by backward selection at a level of 5% with an OLS and a FWLS estimation, respectively.

	Estimate OLS	Estimate FWLS	Std. Error (OLS)	Std. Error (FWLS)
(Intercept)	10.85 ***	11.03 ***	0.16	0.12
AgePainted	-0.004 ***	-0.004 ***	0.001	0.001
Alexander.Calder	-3.24 ***	-3.28 ***	0.17	0.17
Alexej.Jawlensky	-0.38 ***	-0.34 ***	0.09	0.09
Andy.Warhol	-0.75 ***	-0.74 ***	0.07	0.07
Bonhams	-0.42 ***	-	0.19	-
Camille.Pissarro	-0.18 ***	-	0.11	-
Childe.Hassam	-0.58 ***	-0.53 ***	0.16	0.16
Christies	0.19 ***	0.24 ***	0.06	0.07
Collection	0.38 ***	0.34 ***	0.11	0.11
Contemporary	-0.23 ***	-0.17 ***	0.06	0.06
Damien.Hirst	-0.76 ***	-0.68 ***	0.10	0.11
DaySales	-0.28 ***	-0.2 ***	0.05	0.06
Dead	0.68 ***	0.67 ***	0.09	0.09
Donald.Judd	-1.66 ***	-1.53 ***	0.37	0.39
Edgar.Degas	-0.82 ***	-0.78 ***	0.20	0.21
Edouard.Vuillard	-1.10 ***	-1.11 ***	0.11	0.11
Evening	1.37 ***	1.45 ***	0.05	0.05
Georges.Braque	-0.68 ***	-0.68 ***	0.12	0.12
Gerhard.Richter	-0.32 ***	-0.29 ***	0.10	0.10
Hammer	-0.20 ***	-0.27 ***	0.06	0.07
Henri.de.Toulouse.Lautrec	-0.71 ***	-0.65 ***	0.19	0.20
Henri.Matisse	0.47 ***	0.56 ***	0.13	0.14
Henry.Moore	-2.94 ***	-3.07 ***	0.52	0.55
HongKong	1.13 ***	1.17 ***	0.14	0.14
Impressionist	-0.11 ***	-0.16 ***	0.06	0.06
Jean.Michel.Basquiat	-0.61 ***	-0.59 ***	0.10	0.11
Kees.van.Dongen	-0.51 ***	-0.54 ***	0.09	0.09
KollerAuktionen	0.54 ***	0.74 ***	0.20	0.21
London	0.93 ***	0.91 ***	0.07	0.07
Mark.Rothko	0.32 ***	-	0.16	-
Maurice.de.Vlaminck	-1.40 ***	-1.44 ***	0.07	0.07
Max.Ernst	-0.89 ***	-0.85 ***	0.09	0.10
Milan	0.38 **	0.5 **	0.15	0.15
NY	0.87 ***	0.91 ***	0.06	0.07
Pablo.Picasso	0.36 ***	0.39 ***	0.08	0.08
Paris	0.44 ***	0.52 ***	0.07	0.07
Pierre.Auguste.Renoir	-0.43 ***	-0.44 ***	0.07	0.07
Raoul.Dufy	-1.00 ***	-1.02 ***	0.09	0.09
SaintCyr	-0.51 ***	-0.52 ***	0.13	0.14
Sam.Francis	-2.00 ***	-2.06 ***	0.07	0.07
Sothebys	0.13 ***	0.21 ***	0.06	0.06
Surface	-0.00002 ***	-0.00002 ***	0.000001	0.000001
ThemeWord	0.01 ***	0.009 ***	0.002	0.002
Tokyo	-2.79 ***	-	0.23	-
Untitled	-0.44 ***	-0.43 ***	0.06	0.06
VillaGrisebach	0.71 ***	0.83 ***	0.15	0.16
Width	0.02 ***	0.02 ***	0.0007	0.0007
Woman	0.19 ***	0.2 ***	0.06	0.06
Yayoi.Kusama	-1.46 ***	-1.47 ***	0.10	0.11
Heckman Correction	0.10	0.12	0.18	0.20
Adjusted R^2	68%	65%		
Maximum VIF (Variance Inflation Factor)		6.53		
Median VIF (Variance Inflation Factor)		1.46		
Maximum Cook's distance		0.04		
Median Cook's distance		0.0005		
Standard Deviation of Residuals		1.07		

Table 2: Variables and estimators of parameters of probit equation (3) -first stage for Heckman procedure-

	Estimate	Std. error
(Intercept)	0.77 ***	0.04
Contemporary	0.16 **	0.06
Impressionist	-0.15 **	0.07
DaySales	-0.25 ***	0.06
Sam.Francis	-0.14 *	0.08
Kees.van.Dongen	-0.26 **	0.11
Georges.Braque	-0.44 ***	0.14
Edouard.Vuillard	-0.33 **	0.13
Andy.Warhol	-0.42 ***	0.08
Christies	1.04 ***	0.07
Sothebys	0.64 ***	0.06
McFadden Pseudo R2	14%	

at a 5% level, this time using FWLS regression. Table 1 compares results from OLS with those from FWLS. As we expected, the model changes as some variables prove not significantly different from zero at a 5% level with the FWLS estimation. These four variables are Mark Rothko and Camille Pissarro, pieces sold in Tokyo and artworks sold at Bonhams.

The 23 artists (out of 40 available) present in the table have a significant impact on price, ceteris paribus, compared to the other 17 that were not included. However, one should not try to draw a ranking from this table, as difference between artists would not always be statistically significant. Some other results from Table 1 are in line with existing literature: the size (width) has a positive effect on price, but the surface has a negative one, reflecting the idea that a bigger artwork is more expensive, up to the point that it is too big to hang. Prestigious auction houses, like Sotheby's or Christie's are also statistically different from the other ones. Surprisingly, Villa Grisebach (in Germany) stands in the same category. The negative sign linked to the age of the artist reveals that the market prefers, on average, earlier works of the artist whereas untitled artworks are less favoured by the public. Interestingly, mentioning a collection in the title of the sale (for instance: "Important works from the collection of...") leads to higher price. We believe this may be linked to a signal of "good provenance". Also, evening sales tend to exhibit more expensive paintings than day sales.

The second stage of our methodology consists of estimating four continuous time dependent parameters: $\beta(t)$, that will be used to create a price index, $\sigma(t)$, a heteroskedastic term, $\eta(t)$ and $\lambda(t)$ are the parameters that shape the student-skewed distribution of residuals of regression (5). We es-

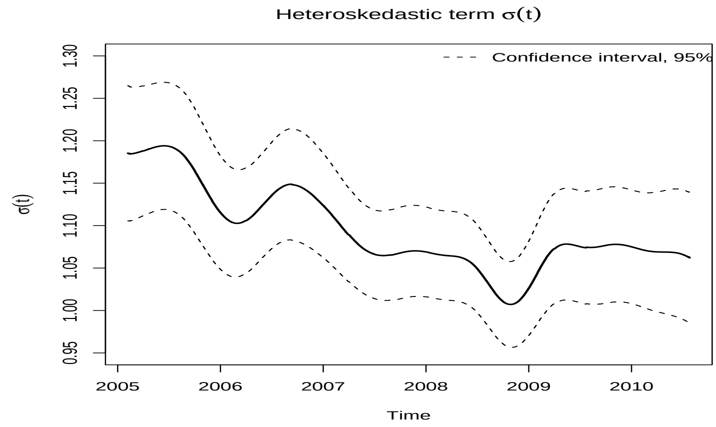

Figure 2: Local maximum likelihood estimator of the heteroskedastic term $\sigma(t)$ from equation (5).

timate numerically the parameters by finding the values that maximize the local log-likelihood function in equation (14).

In order to be as precise as possible, we use the day as unit for t. For the local likelihood estimation, we choose a Gaussian kernel and a bandwidth of $h = 88$ following the plug-in method described above, where the pilot bandwidth h_0 was chosen in the range $h_0 = [1; 30]$.

Figures 2, 3 and 4 plot the estimates of $\sigma(t)$, $\lambda(t)$ and $\eta(t)$, respectively. In order to obtain pointwise confidence intervals, it is necessary to estimate their variance. Figure 8 in appendix illustrates the estimated function $f(t)$ used in equation (16). Practically, this function is estimated by a Nadaraya-Watson estimator.

When considering the parameters, one can first conclude from Figure 3 that we cannot reject the null hypothesis that $\lambda(t) = 0$. In other words, the skewness parameter does not prove useful for this precise example. Nevertheless, we believe one should not draw the conclusion that asymmetry of residuals is typically an unrealistic assumption. For instance, with the same data, we observed that $\lambda(t)$ is significantly different from zero when the Heckman correction is neglected. Concerning the tail parameter $\eta(t)$, it is clear from Figure 4 that tails are fat and that a student distribution better fits data than the Gaussian. For both parameters, however, we cannot conclude that time dependency significantly adds value to the model as compared to

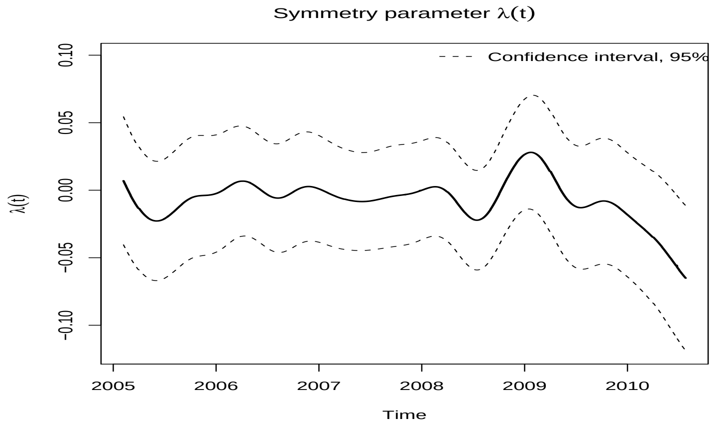

Figure 3: Local maximum likelihood estimator of the symmetry parameter $\lambda(t)$ of equations (5)

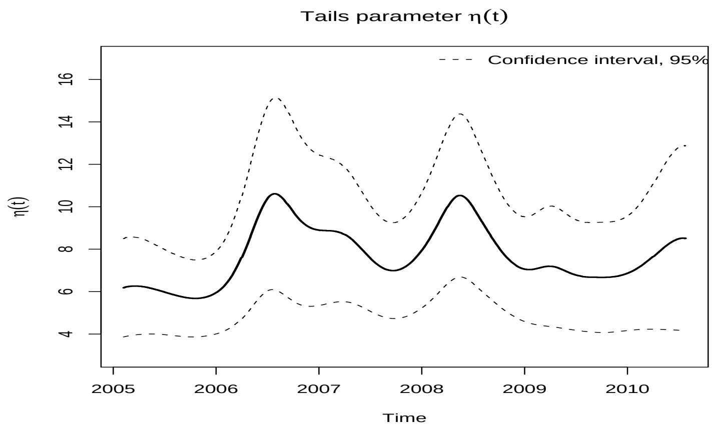

Figure 4: Local maximum likelihood estimator of the degrees of freedom parameter of equations (5)

a constant term.

On the other hand, it is indispensable to allow for heteroskedasticity through a time dependent scaling function. Furthermore, the behaviour of $\sigma(t)$ has an economic meaning: $\sigma(t)$ can be interpreted as the degree of deviation of the realized logged-price of a given painting from the rest of the art market. We call it the "volatility of predictability". In other words, a high $\sigma(t)$ means that is more difficult to predict an artwork's price. A low $\sigma(t)$ corresponds to a more precise estimation of a painting's value. Predictability of prices is vital for auction houses and their clients, especially when guarantees are involved. From Figure 3, we observe that this uncertainty steadily decreased from January 2005 to October 2008. Then, it increased again, or at least stabilized according to the lower confidence interval. It is interesting that the trough of the volatility function occurs at the end of 2008, at about the same time as the peak of the financial crisis with the collapse of Lehman Brothers (September 2008). It also coincides with the drop of the art price index, see Figure 5. This suggests that the precision of the art index has increased during the crisis of 2008/09. An explanation could be the asymmetry of art sales: while there is no upper bound, there is very often a lower bound through a reserve price below which sales are not allowed. Thus, in boom periods there may be a large dispersion due to extreme prices, while in crisis periods, dispersion is smaller since masterpieces are sold at lower values.

The $\beta(t)$ parameters stand for the difference between the returns of a painting cleansed of all its characteristics at a time t, and the average return of this painting through time. This must be compared with the time dummies methodology, where the parameters represent the returns with respect to a given period. We propose a continuous version of Duan (1983)'s and Jones and Zanola (2011)'s smearing estimate. In this framework, a price index whose base value at time $t = 1$ is equal to 100 is given by:

$$Price\ Index(t) = 100e^{\beta(t)-\beta(1)} \times \frac{w_t^{-1} \sum_{i=1}^{N} K(\frac{t_i - t}{h}) \exp(\hat{u}_i)}{w_1^{-1} \sum_{i=1}^{N} K(\frac{t_i - 1}{h}) \exp(\hat{u}_i)}, \tag{19}$$

where $w_t = \sum_{i=1}^{N} K(\frac{t_i - t}{h})$. Note that for the degenerate case $K(\frac{t_i - t}{h}) = d_{i,t}$ we obtain Jones and Zanola (2011) discrete smearing factor. The price index constructed in this way is plotted in Figure 5.

In addition to a daily resolution of time parameters and a higher precision than OLS, we empirically observe that the semi-parametric regression is also less sensitive to lack of data in certain time clusters: as seen in Figure 1,

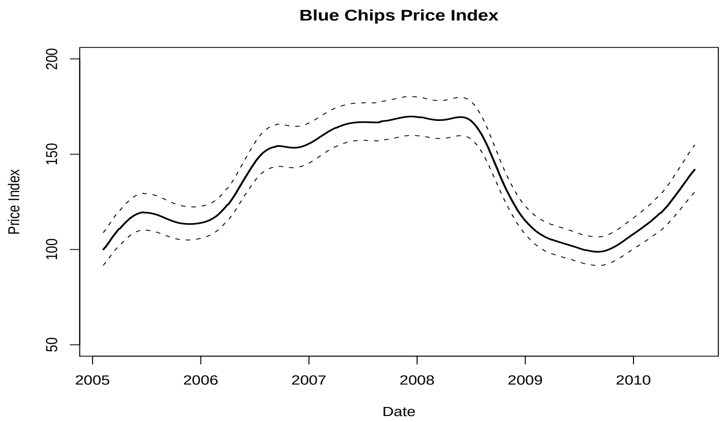

Figure 5: Price index resulting from equation (19): $Price\ Index(t) = 100e^{\beta(t)-\beta(1)} \times S$ where S is a smearing factor and $\beta(t)$ originates from equation (5) and is estimated by maximum likelihood (with local non parametric correction), as shown in equation (13).

the OLS estimation suggests a 87% drop in price in the summer of 2006 and another crash in the summer of 2007. Such impressive drops in prices do not appear in the continuous index in Figure 5. More generally, there is to our knowledge no economic rationale, nor empirical evidence to support the idea that the general level of prices collapsed during the summers of 2006 and 2007. We believe this drop in price shown by the OLS estimation is due to a bias caused by the absence of important sales during summer. Such local flaws are naturally smoothed away by the semi-parametric regression.

3.2 Volatility of index returns

As far as the Blue Chips Index is concerned, it seems an improved methodology based on local maximum likelihood estimation yields more robust results than the traditional OLS methodology. Furthermore, the new regression form presented in equation (5) introduces the concept of volatility of predictability, a measurement that proves useful to better apprehend the discrepancy of valuation of artworks through time.

However, we are also interested in the volatility of the price of a basket of paintings. A possible method to derive volatility of, for instance, quarterly returns when using prior OLS estimation is to consider that the estimated $\beta(t)$ in equation (1) parameters are the "true" observed returns, and compute their volatility, as for any other good quoted in the stock market (see for example Hodgson and Vorkink, 2004). If volatility is assumed constant, then it could be estimated by the sample standard deviation of $\hat{\beta}(t)$, otherwise using e.g. GARCH-type models fitted to the $\hat{\beta}(t)$ process.

Note, however, that the underlying object, $\beta(t)$, is not a stochastic process but rather a deterministic function. It is the expectation of the log-price of a "neutral" painting at time t, and as such does not have a variance. Any attempt to fit time series models designed for stochastic processes to the estimates of $\beta(t)$ is theoretically flawed. What we can do, however, is to assess a degree of variability of this function. Rescaling time as $\tau = t/T$ to map the sample space into the interval $[0, 1]$, the total variation of $\beta(t)$, assuming that $\beta(t)$ is differentiable, is given by

$$TV(\beta) = \int_0^1 |\beta'(\tau)| d\tau,$$

where $\beta'(\tau) = d\beta(\tau)/d\tau$. $TV(\beta)$ is a measure of the overall variability of a function on an interval. On a discretized scale, it could be calculated as the sum of absolute returns, recalling from (19) that log returns over the interval $[t, t+1]$ can be expressed as $\beta(t+1) - \beta(t)$.

Since $TV(\beta)$ is linear in time, it can be further decomposed to obtain, for example, total variations for each year. In our case, we obtain 12.67 % for 2005, 27.90% for 2006, 10.75% for 2007, 59.05% for 2008, and 18.63% for 2009. One can also define $|\beta'(t)|$ as the instantaneous variability of $\beta(t)$, and regard this instantaneous variability as the volatility of the art index, which is time-varying.

Figure 6 plots the estimated instananeous volatility of art along with the VIX index (an index of implied volatility of the S&P 500). Both indices are annualized such that the scales are comparable. The overall level of art and VIX volatilities is about the same, but the art index volatility shows larger swings at the beginning of the sample. One directly observes that, in addition to the change in regime of volatility of predictability as seen previously, the art market suffered from a shock in volatility of prices, linked to a severe drop in returns. This period coincides with the financial crisis in 2008 and the peak in the VIX index. Although the two indices are not directly comparable as the VIX concerns implied volatility whereas our index

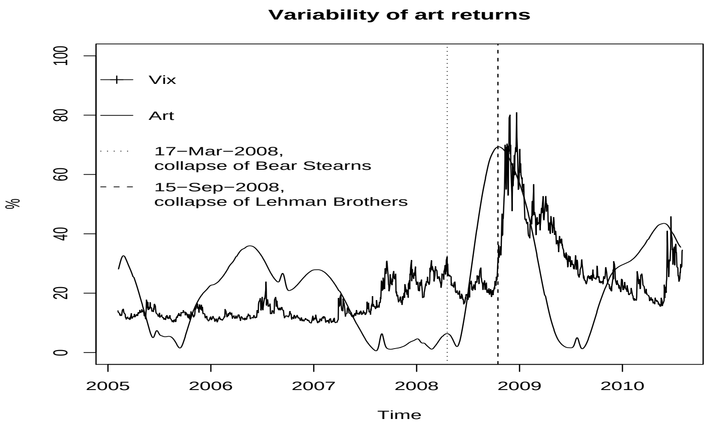

Figure 6: Historical volatility of the art market as compared to implied volatility of S&P 500 options -VIX Index-

concerns instantaneous variability of the index, it seems that the VIX index also suffered from a shock end of 2008, a timing corresponding to Lehman Brothers' bankruptcy.

On the other hand, the apparent high variability of art returns in 2006-2007 is not accompanied by high levels of the VIX. It is our understanding that this variability apparently independent from the stock market was triggered by booming prices of post-war and contemporary art at the time. We believe that the biggest increase in historical volatility of art prices may be linked to the financial crisis, end of 2008. The surge in volatility had serious impact on market participants: in its 2008 third quarter release, Sotheby's affirmed *"These third quarter figures reflect a significant level of losses on our guarantee portfolio principally for fourth quarter sale events including this week's USD10 million Impressionist guarantee losses as well as our estimate of USD17 million on probable guarantee losses in next week's Contemporary sales. We have reduced our guarantee position by 52% as compared to last year and our net guarantee exposure is USD114 million. In this period of considerable economic instability, we will dramatically reduce the guarantees and other special concessions we grant to sellers [...]"*.

Emitting guarantees is equivalent to shorting put options on art. Since the evaluation of such options crucially depends on volatility measures as discussed in this paper, our results may contribute to this new direction of research.

4 Conclusions and outlook

In this paper we have discussed the construction of volatility indices for the art market. In a classical hedonic regression framework, we estimate local parameters, in particular the scale, using a local likelihood approach, which contrasts with the typical OLS estimation method. Our results for a data set comprising blue chip auction data show that the scale parameter is indeed time-varying, which means that the predictability of prices is low when the scale is large, and vice versa. We find that during the financial crisis in 2008/09, this volatility of predictability has been smaller than before, meaning that during this period, price predictions were more precise.

Furthermore, we have considered volatility of the art price index as explained by the variability of the estimated index. We suggest a measure for the degree of variability of the art index and show that for our data, it has about the same magnitude as an implied volatility index on the S&P 500. Art volatility increases similar to the stock index volatility during the financial crisis. Thus, unlike the volatility of predictability, it co-moves with the stock market.

Several applications of these results are possible. For example, to evaluate derivatives on art, such as options, one would have to consider volatility of predictability if the underlying is a single painting, or rather volatility of the art index if the underlying is a large basket or a collection of paintings. For both cases, we have provided suggestions for the evaluation of volatility, but a concise investigation of option pricing on the art market is delegated to future research.

Acknowledgement

We are grateful for helpful comments of participants at the 4th CSDA International Conference on Computational and Financial Econometrics (CFE'10) in London, December 2010. Financial support from the contract *Projet d'Actions de Recherche Concertées* nr. 07/12/002 of the *Communauté francaise de Belgique*, granted by the *Académie universitaire Louvain*, is gratefully acknowledged.

Table 3: Description of data available in the database, per artist. Variables with a “***” are variables whose explanatory power is significant in equation (5) (see Table 1 for more details)

Variable	Description	Number of observations	Proportion
Alexander.Calder ***	Dummy variable: the artist is Alexander Calder (1) or not (0)	41	0.7%
Alexej.Jawlensky ***	id.	159	2.8%
Alfred.Sisley	id.	89	1.6%
Andy.Warhol ***	id.	545	9.7%
Camille.Pissarro ***	id.	110	2.0%
Childe.Hassam ***	id.	52	0.9%
Claude.Monet	id.	143	2.5%
Damien.Hirst ***	id.	328	5.8%
Donald.Judd ***	id.	8	0.1%
Edgar.Degas ***	id.	29	0.5%
Edouard.Vuillard ***	id.	111	2.0%
Edvard.Munch	id.	36	0.6%
Egon.Schiele	id.	17	0.3%
Emil.Nolde	id.	40	0.7%
Ernst.Ludwig.Kirchner	id.	32	0.6%
Georges.Braque ***	id.	90	1.6%
Gerhard.Richter ***	id.	295	5.3%
Henri.de.Toulouse.Lautrec ***	id.	33	0.6%
Henri.Matisse ***	id.	64	1.1%
Henry.Moore ***	id.	4	0.1%
Jean.Michel.Basquiat ***	id.	171	3.0%
Joan.Miro	id.	86	1.5%
Kees.van.Dongen ***	id.	167	3.0%
Lucio.Fontana	id.	172	3.1%
Marc.Chagall	id.	236	4.2%
Mark.Rothko ***	id.	50	0.9%
Maurice.de.Vlaminck ***	id.	325	5.8%
Max.Ernst ***	id.	138	2.5%
Pablo.Picasso ***	id.	222	4.0%
Paul.Gauguin	id.	33	0.6%
Paul.Klee	id.	29	0.5%
Pierre.Auguste.Renoir ***	id.	363	6.5%
Raoul.Dufy ***	id.	167	3.0%
Rene.Magritte	id.	61	1.1%
Richard.Prince	id.	107	1.9%
Sam.Francis ***	id.	482	8.6%
Wassily.Kandinsky	id.	43	0.8%
Willem.de.Kooning	id.	117	2.1%
Yayoi.Kusama ***	id.	225	4.0%
Zao.Wou.Ki	id.	192	3.4%

A Description of data

Table 4: Description of the qualitative data available in the database. Variables with a "***" are variables whose explanatory power is significant in equation (5) (see Table 1 for more details)

Variable	Description	Num. of obs.	Proportion
Dead ***	Dummy variable: the artist is dead (1) or not (0)	4,465	79.6%
DaySales ***	Dummy variable: the auction is a "Day Auction" (1) or not (0)	1,115	19.9%
Morning	Dummy variable: the auction is a "Morning Auction" (1) or not (0)	389	6.9%
Evening	Dummy variable:the auction is an "Evening Auction" (1) or not (0)	1,353	24.1%
Christies ***	Dummy variable: the auction house is Christie's (1) or not	2,198	39.2%
Artcurial	id.	121	2.2%
Bonhams ***	id.	30	0.5%
Dorotheum	id.	17	0.3%
KettererKunst	id.	32	0.6%
KollerAuktionen	id.	29	0.5%
Tokyo	id.	25	0.4%
Phillips	id.	171	3.0%
PierreBerge	id.	9	0.2%
SaintCyr	id.	86	1.5%
Sothebys ***	id.	2,246	40.0%
VillaGrisebach	id.	56	1.0%
Nineteenth	Dummy variable: the auction's theme is based on 19th century art (1) or not (0)	65	1.2%
Collection ***	Dummy variable: the auction's theme is based on a collection (1) or not (0)	114	2.0%
Asian	Dummy variable: the auction's theme is based on Asian art (1) or not (0)	132	2.4%
Contemporary	Dummy variable: the auction's theme is based on contemporary art (1) or not (0)	2,226	39.7%
Impressionist ***	Dummy variable: the auction's theme is based on impressionist art (1) or not (0)	2,134	38.0%
Modern	Dummy variable: the auction's theme is based on modern art (1) or not (0)	2,602	46.4%
PostWar	Dummy variable: the auction's theme is based on post-war art (1) or not (0)	581	10.4%
Surreal	Dummy variable: the auction's theme is based on surrealist art (1) or not (0)	43	0.8%
London ***	Dummy variable: the city where the sales occur is London (1) or not (0)	1,945	34.7%
HongKong ***	id.	111	2.0%
Milan	id.	63	1.1%
NewYork ***	id.	2,196	39.1%
Paris ***	id.	517	9.2%
Monday	Dummy variable: the day of the auction is Monday (1) or not (0)	581	10.4%
Tuesday	id.	1,241	22.1%
Wednesday	id.	1,765	31.5%
Thursday	id.	1,166	20.8%
Friday	id.	470	8.4%
Saturday	id.	200	3.6%
Sunday	id.	189	3.4%
Untitled ***	Dummy variable: the painting's is untitled (1) or not (0)	586	10.4%
Landscape	Dummy variable: the painting's title makes reference to a landscape (1) or not (0)	726	12.9%
Portrait	Dummy variable: the painting's title makes reference to a portrait (1) or not (0)	233	4.2%
StillLife	Dummy variable: the painting's title makes reference to a still life (1) or not (0)	217	3.9%
Animal	Dummy variable: the painting's title makes reference to an animal (1) or not	117	2.1%
Woman ***	Dummy variable: the painting's title makes reference to women (a woman) (1) or not (0)	393	7.0%
Hammer ***	Dummy variable: the price is a hammer price (1), or a premium price (0)	3,223	57.4%

Table 5: Description of time dummy variables

Time dummy	Description	Num. of obs.	Proportion
Y2005Q1	Dummy variable: the quarter of the sale is the first quarter of 2005 (1) or not (0)	150	2.7%
Y2005Q2	id.	420	7.5%
Y2005Q3	id.	42	0.7%
Y2005Q4	id.	320	5.7%
Y2006Q1	id.	175	3.1%
Y2006Q2	id.	519	9.2%
Y2006Q3	id.	25	0.4%
Y2006Q4	id.	396	7.1%
Y2007Q1	id.	234	4.2%
Y2007Q2	id.	539	9.6%
Y2007Q3	id.	38	0.7%
Y2007Q4	id.	463	8.3%
Y2008Q1	id.	248	4.4%
Y2008Q2	id.	426	7.6%
Y2008Q3	id.	229	4.1%
Y2008Q4	id.	274	4.9%
Y2009Q1	id.	141	2.5%
Y2009Q2	id.	348	6.2%
Y2009Q3	id.	35	0.6%
Y2009Q4	id.	313	5.6%
Y2010Q1	id.	151	2.7%
Y2010Q2	id.	126	2.2%

Table 6: Description of the quantitative data available in the database. Variables with a "***" are variables whose explanatory power is significant in equation (5) (see Table 1 for more details)

Variables	Description	Average	Standard Deviation	Min	Max
Height	Height of the painting, in inches	28	21.12	1	195
Width ***	Width of the painting, in inches	28	25.10	1.57	421
Surface ***	The surface of the painting, in inches square				
Lot	Lot Number of the painting	320	321.15	1	7,299
ThemeWord ***	Number of letters for the auction's theme	34	11.51	7	103
WordTitle	Number of letters for the painting's title	20	13.19	2	225
AgePainted ***	The age at which the artist painted the artwork	49	15.99	13	97
AgePainting	The age of the artwork the day of its sale	59	38.44	1	159
Born	The artist's year of birth	1,898	36.73	1831	1,965
Price	The price of the artwork, in USD	1,222,838	3,558,190.72	258	85,000,000
YearPainted	The year the painting was made	1954.77	37.40	1854	2009

B QQ-Plot

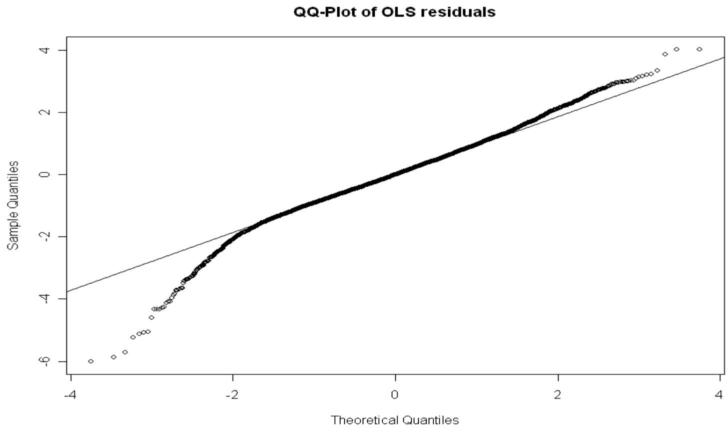

Figure 7: QQ Plot of residuals of regression (1)

C Density of transactions

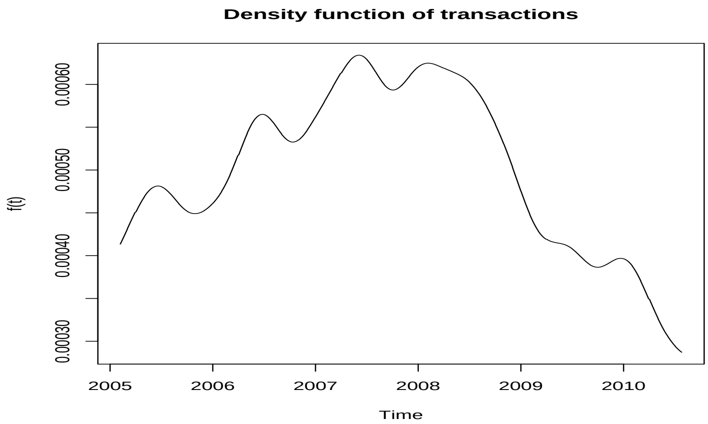

Figure 8: Estimation of the density of transactions $f(t)$ using a Nadaraya Watson estimator

References

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

Bocart, F., Oosterlinck, K., 2011. Discoveries of fakes: Their impact on the art market. Economics Letters 113(2), 124–126.

Boente, G., Fraiman, R., Meloche, J., 1997. Robust plug-in bandwidth estimators in non parametric regression. Journal of Statistical Planning and Inference 57, 109–142.

Campbell, A., Wiehenkamp, C., 2008. Credit default swaps and an application to the art market: a proposal. Open Access publications from Maastricht University urn:nbn:nl:ui:27-19579.

Carroll, R., 1982. Adapting for heteroskedasticity in linear models. Annals of Statistics 10, 1224–1233.

Chanel, O., Gerard-Varet, L., Ginsburgh, V., June 1994. Prices and returns on Paintings: An Exercise on How to Price the Priceless. The GENEVA Papers on Risk and Insurance 19, 7–21.

Collins, A., Scorcu, A., Zanola, R., 2009. Reconsidering hedonic art price indexes. Economics Letters 104(2), 57–60.

Duan, N., 1983. Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association 78, 605–610.

Engle, R., 1982. Estimates of the Variance of U.S. Inflation Based on the ARCH Model. Journal of Money Credit and Banking 15, 286–301.

Engle, R., Rangel, G., 2008. The spline GARCH model for unconditional volatility and its global macroeconomic causes. Review of Financial Studies 21, 1187–1222.

Frey, B., Eichenberger, R., 1995. On the rate of return in the art market: Survey and evaluation. European Economic Review 39, 528–537.

Ginsburgh, V., Mei, J., Moses, M., 2006. On The Computation of Price Indices. Handbook of the economics of art and culture 1, 948–979.

Goetzmann, W., 1993. Accounting for Taste: Art and the Financial Markets over Three Centuries. American Economic Review 83, 1370–1376.

Greenleaf, E., Rao, A. G., Sinha, A. R., 1993. Guarantees in Auctions: The Auction House As Negotiator and Managerial Decision Maker. Management Science 39, 1130–1145.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Hardle, W., 1990. Applied Nonparametric Regression, 1st Edition. Cambridge University Press.

Heckman, J., 1979. Sample Selection Bias as a Specification Error. Econometrica 47, 153–161.

Hendry, D., 2000. Econometrics Alchemy or Science?, 1st Edition. Oxford University Press.

Hodgson, D., Vorkink, K., AUGUST 2004. Asset Pricing Theory and the Valuation of Canadian Paintings. The Canadian Journal of Economics / Revue canadienne d'Economique 37, 629–655.

Jones, A., Zanola, R., 2011. Retransformation bias in the adjacent price index. ACEI S Working Papers Series AWP-01-2011.

McAndrew, C., Thompson, R., 2007. The collateral value of fine art. Journal of Banking and Finance 31, 589–607.

Mei, J., Moses, M., 2002. Art as an investment and the underperformance of master pieces. American Economic Review 92, 1656–1668.

Oosterlinck, K., 2010. The Price of Degenerate Art. Working Papers CEB Université Libre de Bruxelles 09-031.

Pesando, J., 1993. Art as an investment: The Market for Modern Prints. American Economic Review 83, 1075–1089.

Renneboog, L., Spaenjers, C., 2009. Buying beauty: On prices and returns in the art market. Center Discussion Paper Series No. 2009-15.

Rose, R., 1978. Nonparametric estimation of weights in least squares regression analysis, Ph.D. dissertation. University of California at Davis.

Scorcu, A., Zanola, R., 2010. The Right Price for Art Collectibles. A Quantile Hedonic Regression Investigation of Picasso Paintings. Working Paper Series Rimini Centre for Economic Analysis 01_10.

Seckin, A., Atukeren, E., 2006. Art and the Economy: A First Look at the Market for Paintings in Turkey. Economics Bulletin 26, 1–13.

Staniswalis, J., 1989. The Kernel Estimate of a Regression Function in Likelihood-Based Models. Journal of the Statistician American Association 84, 276–283.