*Applied Economics Letters*, 2004, **11**, 245–249

# *Do art specialists form unbiased pre-sale estimates? An application for Picasso paintings*

CORINNA CZUJACK and MARIA FRAGA O. MARTINS*‡

*ECARES-ULB, Av. F. Roosevelt 50, CP 177, B-1050 Bruxelles, Belgium* and
*‡ISEGI-UNL, Campus de Campolide, 1070-124, Lisboa, Portugal*

This work investigates whether art specialists provide good predictors of realized prices for Picasso paintings. A sample selection model is proposed to represent the decision of the seller and the price equation. The model is applied to data on 675 Picasso paintings for the period 1975–1994. It is found that the two auction houses, Sotheby's and Christie's, have given good predictions for the works that have been sold. However, for the unsold works, it would have been possible to give estimates better than those given by the salerooms. As a consequence they could perhaps have sold more paintings than they actually did.

## I. INTRODUCTION

In the period 1975 to 1994 about 675 Picasso paintings have been offered to the market by Sotheby's and Christie's, of which 499 were sold. The unbiasedness of pre-sale estimates is tested for. The study is interested in whether art specialists provide good predictors of realized prices and whether there are differences between both houses. It is argued that pre-sale estimates should be unbiased because, to attract a seller, the auction house cannot set the pre-sale estimate too low. The seller has little influence on the pre-sale estimate, but he sets the reservation price under which he is not willing to sell. The auction house can choose to set the pre-sale estimate range in a way that the client offers the work for sale there. A potential buyer, on the other hand, knowing this relationship, will not attend an auction when the pre-sale estimate is too high. As 26% of the works remained unsold – that is, were over-estimated – prices are predicted based on the information available to see whether we could have done better. The outline of the paper is as follows. Section II describes the basic econometric model. Section III reports the main empirical results and tests for unbiasedness. Section IV concludes.

## II. ECONOMETRIC MODEL

The study tests whether pre-sale estimates are unbiased predictions of realized prices as follows. Let $X_i$ be the pre-sale estimate of the price $Y_i$ of a lot $i$. Unbiasedness requires the best estimate to be equal to $X_i = E(y_i|\Omega)$, where $\Omega$ is the information set available to the expert. The hammer price, $Y_i$, will be given by,

$$Y_i = X_i + u_i \qquad (1)$$

where $u_i$ is a random disturbance with zero mean and $E(u_i|\Omega)$ should be equal to zero, implying that the estimate $X_i$ takes into account all the information contained in $\Omega$. To test whether $X_i$ is an unbiased estimator of $Y_i$, one could run regression 2 by ordinary least squares:

$$Y_i = \beta_0 + X_i'\beta_1 + u_i \qquad (2)$$

and test the hypothesis of $\alpha = 0$ and $\beta = 1$. If this hypothesis is not rejected then $X_i$ is an unbiased estimator of $Y_i$. However, in the framework, the dependent variable $Y_i$ is truncated, i.e. the price exists only for the sold works, even if there is full information on $X$, the pre-sale estimate. The truncation of the dependent variable is based not on the

* Corresponding author. E-mail: mrfom@isegi.unl.pt

*Applied Economics Letters* ISSN 1350–4851 print/ISSN 1466–4291 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13504850410001674894

value of the dependent variable, i.e. the hammer price, but rather on the value of another variable that is correlated with it; in this case this is the reservation price. If the hammer price exceeds the reservation price, the work is sold. Otherwise the work is unsold. Formally, the model can be written as a sample selection model, as follows:

$$Z_i = 1(W_i'\gamma + \varepsilon_i > 0), \quad \text{decision process} \tag{3}$$

$$Y^*{}_i Z_i = Y_i = (X_i'\beta + u_i)Z_i, \quad \text{price equation.} \tag{4}$$

Equation 3 formulates the decision of the seller to sell or not to sell. The relationship between the price and the pre-sale estimate described in Equation 4 is called, for simplicity, the price equation. $Y_i$ is the value of lot $i$ (the hammer price if the work is sold and zero otherwise), $Y_i^*$ the price paid for lot $i$, $Z_i$ is a dummy variable that is equal to one if it is sold and zero otherwise, $n_1$ is the number of paintings sold and $n$ represents the total number of works offered to the market. $W$ contains the variables influencing the selling decision, i.e. the variables that influence the hammer price and the variables which determine the reservation price; the function 1(.) is the usual indicator function. The estimation of the price equation (4) for only those works which are sold could result in biased estimates since, in general, $E(Y_i^*|Z_i = 1, W_i) \neq X_i'\beta$. The standard practice is to correct this sample selection bias.

If it is assumed that $(u_i, \varepsilon_i)$ in Equations 3 and 4 are bivariate normally distributed, independently of $(W_i, X_i)$ with zero mean and $E(u_i^2) = \sigma, E(\varepsilon_i^2) = 1$, $\text{cov}(u_i, \varepsilon_i) = \rho\sigma$, the parameters can be estimated by maximum likelihood or by the two-step Heckman (1974, 1979)[1] estimator. The latter is based on the fact that the price equation can be rewritten as:

$$E(Y_i^*|Z_i = 1, W_i) = X_i'\beta + \rho\sigma\lambda(W_i'\gamma) \tag{5}$$

where $\lambda$ is the inverse Mills ratio defined as $\lambda(.) = \varphi(.)/\Phi(.)$, with $\varphi$ and $\Phi$ being the standard normal density and distribution functions.

Heckman's method works as follows: In the first step, an ordinary Probit model is used to obtain consistent estimates $\hat{\gamma}$ of the parameters of the decision equation. In the second step, the selectivity regressor $\lambda$ is evaluated at $\hat{\gamma}$ and regression 5 is estimated by OLS for the observations with $Y_i > 0$. This regression provides a test for sample selectivity, as well as an estimation technique. The coefficient on the selectivity regressor is $\rho\sigma$. Since $\sigma \neq 0$, the ordinary $t$ statistic for this coefficient to be zero can be used to test the hypothesis that $\rho = 0$. If this coefficient is not significantly different from zero, it may be decided that selectivity is not a problem for this data set and proceed to use least squares as usual. These estimators are consistent, but inefficient. However, they provide good starting values for maximum likelihood (ML) estimation. Under the assumptions mentioned above, ML estimators are efficient and can be obtained by maximizing $L$:

$$L = \prod_{Z=0} \Phi[-W_i'\gamma] \prod_{Z=1} \left[\frac{1}{\sigma_u}\phi\left\{\frac{(Y_i - X_i'\beta)}{\sigma_u}\right\}\right] \times \Phi\left[\left\{W_i'\gamma + \rho\left(\frac{Y_i - X_i'\beta}{\sigma_u}\right)\right\}(1-\rho^2)^{-\frac{1}{2}}\right] \tag{6}$$

## III. EMPIRICAL RESULTS

### *The data*

The data set used consists of all paintings sold by Sotheyby's and Christie's between 1975 and 1994. For each painting information has been collected on the work itself, the sale and on the agents who traded it. Sources for the construction of the database are Mayer's (1964 to 1996) Compendia 'Annuaire de Ventes' and the pre-sale catalogues published by the auction houses. The following information is included: the title of the painting, place and date of sale, dimensions, date of creation (and therefore, working periods: cubism, surrealism, etc.), technique and medium used, price and pre-sale estimate and information about exhibitions at which the work was shown. The database further contains the Zervos Catalogue Raisonné number, information on the signature and on whether it has been illustrated in the catalogue.

During the observed period, Sotheby's sold twice as many Picasso paintings as Christie's but also had twice as many of the artist's works that went unsold. The percentage of unsold works is roughly 26% for both houses (Table 1).

Table 2 compares the price $p$ with the pre-sale estimated range $[\hat{p}_{MIN}, \hat{p}_{MAX}]$ for each work offered by Sotheby's and Christie's. It is found that both houses overestimated nearly 50% of all paintings, 30% of works were sold at a price within the estimate range and 20% were

Table 1. *Number and shares of sold and unsold works*

| | Number | | | Percentage | | |
|---|---|---|---|---|---|---|
| | Sold works | Unsold works | Total | Sold works | Unsold works | Total |
| Christie's | 182 | 65 | 247 | 73.68 | 26.32 | 100 |
| Sotheby's | 317 | 111 | 428 | 74.07 | 25.93 | 100 |
| Total | 499 | 176 | 675 | 73.93 | 26.07 | 100 |

[1] In most applications (labour market) this model is used with the intention to test for sample selectivity bias, (in our case, whether it matters to include the unsold works). It is considered rather as the estimation procedure to use when having the intention to include the decision process (sold only when hammer price is higher than reservation price).

Table 2. *Observed under- and overestimation*

| | Percentage* | | | | Number | |
|---|---|---|---|---|---|---|
| | | Sold | | | | |
| | Unsold | $p < \hat{p}_{MIN}$ | $\hat{p}_{MIN} \leqslant p \leqslant \hat{p}_{MAX}$ | $p < \hat{p}_{MAX}$ | Unsold | Sold |
| Sotheby's | | | | | | |
| 1975–1980 | 26.22 | 12.03 | 22.78 | 38.97 | 19 | 54 |
| 1981–1986 | 23.53 | 21.80 | 40.93 | 13.73 | 32 | 104 |
| 1987–1990 | 28.48 | 16.40 | 22.96 | 32.16 | 45 | 113 |
| 1991–1994 | 24.59 | 37.65 | 23.19 | 14.57 | 15 | 46 |
| Total | 25.70 | 21.97 | 27.47 | 24.86 | 111 | 317 |
| Christie's | | | | | | |
| 1976–1980 | 3.33 | 38.19 | 30.38 | 28.10 | 1 | 29 |
| 1981–1986 | 28.07 | 26.56 | 37.86 | 7.50 | 16 | 41 |
| 1987–1990 | 34.00 | 11.83 | 31.77 | 22.41 | 34 | 66 |
| 1991–1994 | 30.43 | 22.88 | 22.26 | 24.43 | 14 | 46 |
| Total | 23.96 | 24.86 | 30.57 | 20.61 | 65 | 182 |

*Values sum up to 100 horizontally.

underestimated. In particular for Sotheby's an increase is found of strong overestimation over the years. Because of the high number of unsold works, it is reasonable to test whether art specialists give unbiased estimators.

### Estimation results

As pointed out before, is the price $Y_i$ is assumed to be determined only the pre-sale estimate, $X_i$. This formulation is justified by arguing that the pre-sale estimate is influenced by the same variables that influence the hammer price,[2] because the experts have the same information available as the market has. Pre-sale estimate is taken as the midpoint of the range of the pre-sale estimate.[3] The variables $W$, the regressors for the decision equation, thus include the pre-sale estimate, dimensions, dummies for working periods of the artist, the number of times the painting has been exhibited, years of sale, techniques and media used and whether a work has been published on Zervos' catalogue or not. The variable $Z$ is equal to 1 if the paining has been sold, and 0 otherwise. For the specific application, one distinguishes between the auction houses Sotheby's and Christie's because it is suspected they may behave differently. Thus Equation 2 is written as

$$Y_i = \beta_0 + \beta_{0s}\delta_{si} + X_i'\beta_1 + X_i'\beta_{1s}\delta_{si} + u_i \qquad (7)$$

where $\delta_{si} = 1$ if lot $i$ was sold at Sotheby's and $\delta_{si} = 0$ if sold at Christie's.

Table 3 shows estimation results for the proposed model (sample selection model). It is estimated by Heckman's two step method and by maximum likelihood. For comparison, the case in which there is no correlation ($\rho = 0$) is also considered. This yields OLS estimates for the price equation and the Probit maximum likelihood estimates for the decision equation.

It is found that the coefficients and standard errors do not differ much across the different approaches. Results obtained from the Heckman two-step estimator show a $\lambda$ that is not significantly different from zero, suggesting that no selection bias from including only the works sold is present. This is equivalent to not being able to reject the hypothesis that $\rho = 0$ (no correlation). Since the Heckman estimator is not efficient the maximum likelihood results are also analysed. The hypothesis of $\rho = 0$ cannot be rejected at the 5% significance level, also suggesting that selectivity is not a problem. Consequently, one should be able to continue applying ordinary least squares.

### Testing for unbiasedness and prediction

The purpose of this study is to test whether pre-sale estimates are unbiased predictors of realized prices. Three different hypotheses are tested. First, the null hypothesis that both houses behave identically in predicting prices is tested. Then the study tests whether each house releases unbiased predictions. Finally, the joint hypothesis that Sotheby's and Christie's behave identically is tested and give unbiased pre-sale estimates. Results are displayed in Table 4. None of the hypotheses can be rejected.

These results suggest that the auction houses have given good predictions for the works that have been sold. A work of art is bought in, if it has been overestimated, i.e. if no one has been willing to pay at least the reservation price. One wonders whether this overestimation of prices by Sotheby's or Christie's is due to the fact that the market is determined by forces that specialists have not been able

[2] A regression has been run regressing the pre-sale estimate on the same regressors which confirms this.
[3] This is, of course, a strong simplification.

Table 3. *Results for the parametric estimations of the sample selection model*

| | Heckman two-step method | | ML method | | $\rho=0$ | |
|---|---|---|---|---|---|---|
| | Coeff. | S.E. | Coeff. | S.E. | Coeff. | S.E. |
| *Decision equation* | | | | | | |
| Periods | | | | | | |
| Childhood and youth | −0.609 | 0.955 | −0.509 | 1.005 | −0.609 | 0.955 |
| Blue and Rose | 0.00 | – | 0.00 | – | 0.00 | – |
| Cubism | −1.429 | 0.914 | −1.490 | 1.018 | −1.429 | 0.914 |
| Camera and classicism | −1.169 | 0.892 | −1.224 | 0.939 | −1.169 | 0.892 |
| Juggler of the form | −0.999 | 0.891 | −1.081 | 0.920 | −0.999 | 0.891 |
| Guernica | −0.814 | 0.884 | −0.887 | 0.926 | −0.814 | 0.884 |
| Politics | −1.132 | 0.889 | −1.173 | 0.921 | −1.132 | 0.889 |
| The old Picasso | −1.594 | 0.874 | −1.619 | 0.903 | −1.594 | 0.874 |
| Techniques | | | | | | |
| Oil | 0.00 | – | 0.00 | – | 0.00 | – |
| Collage | −0.946 | 0.217 | −0.821 | 0.347 | −0.946 | 0.217 |
| Mixed media | 0.305 | 0.576 | 0.279 | 0.744 | 0.305 | 0.576 |
| Media | | | | | | |
| Canvas | 0.00 | – | 0.00 | – | 0.00 | – |
| Cardboard | 0.322 | 0.406 | 0.412 | 0.543 | 0.322 | 0.406 |
| Panel | −0.218 | 0.613 | −0.355 | −1.016 | −0.218 | 0.613 |
| Paper | 1.143 | 0.628 | 1.126 | 0.606 | 1.143 | 0.628 |
| Published by Zervos | 2.546 | 0.198 | 2.514 | 0.246 | 2.546 | 0.198 |
| Exhibitions | | | | | | |
| Not exhibited | 0.00 | – | 0.00 | – | 0.00 | – |
| Exhibited 1 to 5 times | −0.425 | 0.163 | −0.423 | 0.189 | −0.425 | 0.163 |
| Exhibited more than 5 times | −0.428 | 0.428 | −0.391 | 0.530 | −0.428 | 0.428 |
| Time dummies | 18 var.* | | 18 var.* | | 18 var.* | |
| Constant $\beta_0$ | 2.333 | 0.886 | 2.367 | 1.038 | 2.333 | 0.886 |
| SO's dummy $\beta_{0s}\ \delta_{si}$ | −0.351 | 0.202 | −0.365 | 0.241 | −0.351 | 0.202 |
| Pre-sale estimate $\beta_1$ | −0.013 | 0.012 | −0.010 | 0.016 | −0.013 | 0.012 |
| Pre-sale estimate* SO's dummy $\beta_{1s}\ \delta_{si}$ | 0.022 | 0.015 | 0.024 | 0.019 | 0.022 | 0.015 |
| $\rho$ | | | −0.270 | 0.230 | 0.00 | |
| *Price equation* | | | | | | |
| Constant $\beta_0$ | −0.801 | 1.268 | −0.310 | 3.200 | −0.999 | 2.223 |
| SO's dummy $\beta_{0s}\ \delta_{si}$ | −0.184 | 1.493 | 0.006 | 3.600 | −0.264 | 2.833 |
| Pre-sale estimate $\beta_1$ | 1.308 | 0.083 | 1.309 | 0.004 | 1.300 | 0.409 |
| Pre-sale estimate* SO's dummy $\beta_{1s}\ \delta_{si}$ | 0.007 | 0.090 | 0.002 | 0.004 | 0.009 | 0.500 |
| $\lambda$ | −1.104 | 1.880 | | | | |
| $R^2$ | 0.759 | | 0.759 | | 0.759 | |
| Log $L$ | −1960.8 | | −2155.9 | | −1963.5 | |

*Because of the large number, coefficients are not given here.

to measure, or whether specialists did not predict correctly although the available information would have allowed them to do so. The data are used to test whether better estimates could have been given.

Art specialists, when determining pre-sale estimates, have only information on sales that have taken place in the past. For this reason OLS (hedonic) regressions are run for each year for which one wants to predict by including only the previous years' data.[4] The predictions $\hat{p}$ are then compared with the specialists' predictions $[\hat{p}_{MIN}, \hat{p}_{MAX}]$. As a result, 36% of predictions lie below the specialists' low estimates, 38% within the specialists' estimate range and 26% above their maximum estimate. The large share of 36%, for which the predictions lie below the specialists' forecasts, suggests that it would have been possible to give estimates closer to the true value, with the consequence that more paintings might have been sold.

## IV. CONCLUSIONS

The study has tested whether the art specialists at Sotheby's and Christie's have given good predictors of realized prices.

[4] Specifically, hedonic regressions of the log of price were ran on the following variables: dimensions, techniques and media used, exhibitions, signature, Zervos publication, working periods, a dummy for resales, provenance, place of sale, time dummies and the log (of the midpoint) of the pre-sale estimate.

Table 4. *Testing for unbiasedness; Wald test*

| Hypothesis | $\chi^2$-critical OLS | Value | Result |
|---|---|---|---|
| Both houses identical $\beta_{0s}=0,\ \beta_{1s}=0$ | 0.122 | 5.99 | accepted |
| Unbiasedness of Christie's $\beta_0=0,\ \beta_1=1$ | 2.685 | 5.99 | accepted |
| Unbiasedness of Sotheby's | 3.743 | 5.99 | accepted |
| $\beta_0+\beta_{0s}=0,\ \beta_1+\beta_{1s}=1$ | 3.743 | 5.99 | accepted |
| Both houses identical and unbiased $\beta_{0=}\beta_{0s}=0,\ \beta_{1=1}\,\beta_{1s}=0$ | 6.428 | 9.49 | accepted |

This was done by applying a sample selection model that allowed one to take into account the decision process of a seller to sell a work only if the hammer price exceeds the reservation price. Since there was no selectivity bias from using only the works sold, one was able to continue applying ordinary least squares. Four hypotheses were tested giving the results that both houses behave identically and have accurately predicted prices for the works sold. Then the 26% of unsold works, which were overestimated, were taken and prices predicted for them and compared these with the experts' estimates. The results suggest that, with the information available, it would have been possible to give estimates better than those given by the salerooms. As a consequence they could perhaps have sold more works than they actually did.

## ACKNOWLEDGEMENTS

We are grateful to R. Flôres and V. Ginsburgh for useful comments.

## REFERENCES

Heckman J. (1974) Shadow prices, market wages and labour supply, *Econometrica*, **42**, 679–94.

Heckman J. (1979) Sample selection as a specification error, *Econometrica*, **47**, 153–61.