

HOW WELL DO SELECTION MODELS PERFORM? ASSESSING THE ACURACY OF ART AUCTION PRE-SALE ESTIMATES

Binbing Yu and Joseph L. Gastwirth

National Institute on Aging and The George Washington University

Abstract: Art auction catalogs provide a pre-sale prediction interval for the price each item is expected to fetch. When the owner consigns art work to the auction house, a reserve price is agreed upon, which is not announced to the bidders. If the highest bid does not reach it, the item is brought in. Since only the prices of the sold items are published, analysts only have a biased sample to examine due to the selective sale process. Relying on the published data leads to underestimating the forecast error of the pre-sale estimates. However, we were able to obtain several art auction catalogs with the highest bids for the unsold items as well as those of the sold items. With these data we were able to evaluate the accuracy of the predictions of the sale prices or highest bids for all item obtained from the original Heckman selection model that assumed normal error distributions as well as those derived from an alternative model using the t_2 distribution, which yielded a noticeably better fit to several sets of auction data. The measures of prediction accuracy are of more than academic interest as they are used by auction participants to guide their bidding or selling strategy, and similar appraisals are accepted by the US Internal Revenue Services to justify the deductions for charitable contributions donors make on their tax returns.

Key words and phrases: Art auction, forecast error, nonignorable missing data, selection model.

1. Introduction

The selection model was introduced by Heckman (1976) to describe the distribution of wages that women could earn if they participated in the labor force. As data on market wages is only available for working women, whose market wage exceeds the value of their work at home, the observed wages are a selected sample from the potential wages distribution of women. The use of selection models is the subject of considerable discussion in the econometrics and statistics literature, much of it focusing on the issues of identifiability and sensitivity (Kenward (1998)). Since the true values of missing data are rarely obtained it is difficult to evaluate how well the adjusted estimates obtained from a selection model actually fit the “full” data set. The accidental finding of two art auction

catalogs with both the sale price for the sold times and the highest bids received for the unsold items in a used bookstore by the second author stimulated this study. These data sets enable us to estimate the accuracy of the pre-sale predictions and to evaluate the selection models more thoroughly than in the previous literature.

In this article, we extend Heckman's selection model to a heavier tailed error distribution and demonstrate that this model yields more accurate prediction of the "missing" highest bids using real data and simulation. The proposed method is parametric so it is applicable when the selection and response equations have the same predictors, while other semi-parametric approaches require that there is at least one independent variable in one of the equations that is not used in the other (Vella (1998)). The result that the selection bias leads to underestimation of the forecast error of the pre-sale estimates has implications for participants in the art market and in the use of appraisals for the purpose of deducting a charitable deductions or the valuation of art in legal matters (Inde (1998, Chap. 4)). Prospective bidders need an accurate estimate of the forecast error in order to plan their bidding strategy.

The art auction data and measures of forecast error are described in Section 2. The effect of selection bias is illustrated by comparing the estimated forecast error measures when the data on unsold items are omitted to those calculated from the full data. The use of Heckman's model and its extension are described and applied to the art auction data in Section 3. Criteria for evaluating the utility and fit of selection models are given in Section 4. The results using the extended selection model are shown to be much closer to the true data than those obtained from the original selection model, which assumes the errors are normally distributed.

2. Art Auction Data and The Selection Bias

In art auctions, all items offered are subject to a reserve price that is not announced to the potential bidders. The auction catalogs provide an interval prediction (L, U) for the price, and the middle of the interval is considered as the predicted price (P) . Both the reserve price (R) and highest bid (A) should be related to the predicted price (P) . For evaluating the accuracy of the pre-sale prediction P published by the auction houses, several measures of forecast error have been used (Gastwirth (1979)). Prospective buyers are interested in the difference between the highest bid (sale price for a sold item) and the predicted price, expressed as a percent error $|A - P|/A$, as they need to know how much money beyond the published estimate they may need to bid in order to obtain the item. On the other hand, $|A - P|/P$ may be of more interest to sellers who

Table 2.1. Summary of the auction catalogs.

| Auction House | Number | Year | Description | # of items | # of sold items |
|---------------|--------|------|-------------------------------------|------------|-----------------|
| Sotheby's | 3,850 | 1998 | Jewels | 508 | 342 |
| Sotheby's | 6,371 | 1992 | Chinese Paintings | 186 | 110 |
| Christie's | 8,990 | 1998 | 19 & 20th Cent. Prints | 729 | 617 |
| Christie's | 9,028 | 1998 | 20th Cent. Contemporary Prints | 205 | 157 |
| Christie's | 9,038 | 1999 | 19 & 20th Cent. Old Master's Prints | 595 | 512 |

are concerned with their expected revenue from the sale. The two corresponding measures are the average error relative to the actual price,

$$AREA = n^{-1} \sum_{i=1}^n \frac{|A_i - P_i|}{A_i},$$

and the average error relative to the predicted price,

$$AREP = n^{-1} \sum_{i=1}^n \frac{|A_i - P_i|}{P_i},$$

where n is the total number of items that actually were offered at the sale.

When the highest bid exceeds or equals the reserve price, i.e., $A \geq R$, the item is sold and the sale price is the highest bid. When the highest bid fails to reach the reserve price, the item is not sold and that final bid A is omitted from the published post-sale report. Since the sale of an item at the auction depends on both the reserve price and the highest bid received, the published sale price data are a non-random sample of the final bids. Selection bias occurs because the highest bid A is missing when $A < R$. Consequently, estimates of the AREA and AREP based only on the sale prices of the sold items are biased.

We obtained several catalogs along with all the highest bids, including those for unsold items, enabling us to evaluate the selection bias due to the unsold items, whose highest bids are usually not released by the auction houses. Table 2.1 summarizes the year, type of art, and number of objects for sale in each catalog. As reported in the supplementary document, the frequencies of the highest bid falling below, within and above the predicted interval, even when the intervals are expanded by a factor of 1.75, show that the auctioneers underestimate the variability of the actual bids.

In Table 2.2, we present the forecast error measures *AREA* and *AREP* and the associated standard errors calculated on the highest actual bids for the sold items, the unsold items, and the entire auction sale. From Table 2.2, we can see that both average relative errors for the sold items are lower than those for the complete data, and much smaller than those for the unsold items. Therefore,

Table 2.2. Estimates and the standard errors (SE) of measures of forecast error by sale status.

| Auction | Items used | Number | $AREA$ | $SE(AREA)$ | $AREP$ | $SE(AREP)$ |
|---------|------------|--------|--------|------------|--------|------------|
| 3,850 | All | 508 | 0.325 | 0.011 | 0.295 | 0.012 |
| | Sold | 342 | 0.224 | 0.008 | 0.277 | 0.017 |
| | Unsold | 166 | 0.534 | 0.023 | 0.331 | 0.008 |
| 6,371 | All | 186 | 0.577 | 0.034 | 0.437 | 0.041 |
| | Sold | 110 | 0.339 | 0.024 | 0.426 | 0.068 |
| | Unsold | 76 | 0.921 | 0.055 | 0.453 | 0.014 |
| 8,990 | All | 729 | 0.353 | 0.010 | 0.327 | 0.014 |
| | Sold | 617 | 0.290 | 0.009 | 0.315 | 0.016 |
| | Unsold | 112 | 0.698 | 0.030 | 0.394 | 0.009 |
| 9,028 | All | 205 | 0.395 | 0.024 | 0.307 | 0.019 |
| | Sold | 157 | 0.288 | 0.017 | 0.276 | 0.025 |
| | Unsold | 48 | 0.745 | 0.061 | 0.406 | 0.014 |
| 9,038 | All | 595 | 0.329 | 0.010 | 0.366 | 0.023 |
| | Sold | 512 | 0.280 | 0.010 | 0.364 | 0.026 |
| | Unsold | 83 | 0.628 | 0.026 | 0.374 | 0.009 |

calculating the relative error using only the sold items underestimates the forecast error of the pre-sale predictions. In this paper, the publicly released data, which includes all the predicted prices, sale status, and the highest bids for the *sold* items, is referred to as available *incomplete* data. The *full* data includes the publicly released data plus the highest bids for the unsold items. As the reserve prices are confidential, we could not obtain them so the full data is still not complete.

As the results in Table 2.2 indicate, the sale prices are right skewed, we fit a linear regression $\log(A_i) = \beta_0 + \beta_1 \log(P_i) + \sigma\epsilon_i$, $\epsilon_i \sim N(0, 1)$, $i = 1, \dots, n$, to the full data as well as the sold and unsold items to examine the selection bias. The parameter estimates and the associated standard errors are shown in Table 2.3. The range of the standard errors for all the intercepts is (0.13, 0.45) and the range of the standard errors of all the slopes is (0.02, 0.05). Only the slopes β_1 for all items in Sale 3,850 was statistically different from 1. The differences of the intercepts β_0 between the sold and unsold items are pronounced. If we assume common slopes for the sold and unsold items, the intercepts for the sold and unsold items are statistically different for all five auction sales (p -value < 0.0001). The regression lines for Sale 8,990, which are typical, are shown in Figure 2.1. The intercept of the regression for the unsold items (indicated by o) is much lower than that of the sold items (indicated by *), while the regression line for all sold items lies above the line for unsold items. The regression line based on all items is between the lines for the sold and unsold items. It is generally true

Table 2.3. Parameter estimates and the associated standard error (SE) from the regression by sale status.

| Auction | Items used | $\hat{\beta}_0$ | SE($\hat{\beta}_0$) | $\hat{\beta}_1$ | SE($\hat{\beta}_1$) | $\hat{\sigma}$ | R^2 |
|---------|------------|-----------------|-----------------------|-----------------|-----------------------|----------------|-------|
| 3,850 | All | 0.431 | 0.137 | 0.919 | 0.021 | 0.343 | 0.795 |
| | Sold | 0.343 | 0.149 | 0.955 | 0.023 | 0.304 | 0.835 |
| | Unsold | -0.469 | 0.128 | 1.009 | 0.019 | 0.178 | 0.946 |
| 6,371 | All | -0.083 | 0.363 | 0.979 | 0.042 | 0.496 | 0.737 |
| | Sold | -0.001 | 0.446 | 0.996 | 0.052 | 0.464 | 0.775 |
| | Unsold | -0.572 | 0.282 | 0.999 | 0.032 | 0.325 | 0.868 |
| 8,990 | All | 0.136 | 0.136 | 0.974 | 0.016 | 0.386 | 0.839 |
| | Sold | 0.185 | 0.140 | 0.977 | 0.016 | 0.363 | 0.853 |
| | Unsold | -0.845 | 0.146 | 1.038 | 0.017 | 0.167 | 0.972 |
| 9,028 | All | -0.045 | 0.205 | 0.984 | 0.022 | 0.358 | 0.909 |
| | Sold | -0.073 | 0.215 | 0.998 | 0.023 | 0.332 | 0.923 |
| | Unsold | -0.592 | 0.236 | 1.006 | 0.025 | 0.191 | 0.973 |
| 9,038 | All | 0.003 | 0.150 | 0.998 | 0.017 | 0.414 | 0.848 |
| | Sold | 0.061 | 0.153 | 1.000 | 0.018 | 0.393 | 0.862 |
| | Unsold | -0.453 | 0.137 | 0.999 | 0.016 | 0.202 | 0.958 |

that the regression lines for sold items are above those for unsold items. This is reasonable as the highest bids for unsold items are below their reserve prices, which are often a percentage, e.g., 60-80%, of the predicted prices. Thus, the intercept term is negative in the regression relating $\log(A)$ and $\log(P)$. For all five auctions the R^2 values are higher and $\hat{\sigma}$ values are smaller in the regressions for unsold items than those of the regressions for sold items. The lower regression lines, larger R^2 , and smaller $\hat{\sigma}$ for the unsold items means that the final bids for the unsold times are substantially lower than their predicted prices, but have less variation. This implies that there is more variation for the highest bid for the sold items. This could be due to extremely high bids (outliers) for some items and/or lower reserve prices for others.

3. Selection Models for Art Auction Data

Heckman's model concerns an incompletely observed variable Y that has a linear regression on covariates \mathbf{X} and is observed if and only if the value of another completely unobserved variable D exceeds a threshold (say zero). The distribution $f(Y_i, D_i | \mathbf{X}_i; \theta)$ is specified as follows:

$$\begin{pmatrix} Y_i \\ D_i \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \beta^T \mathbf{X}_i \\ \xi^T \mathbf{X}_i \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right], \quad (3.1)$$

where \mathbf{X} denotes the covariates, Y is incompletely observed, D is never observed. The regression coefficients are β and ξ , and $N_2(a, b)$ denotes the bivariate normal

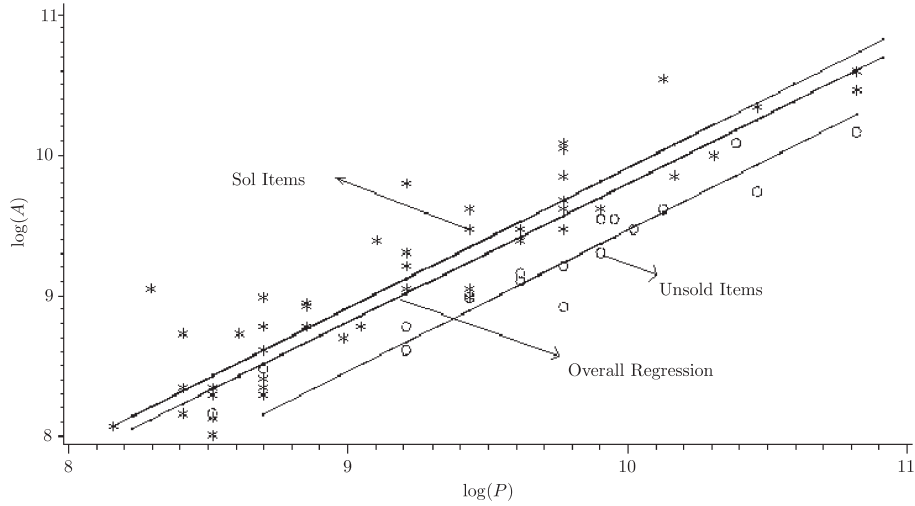


Figure 2.1. Plots of $\{\log(P), \log(A)\}$ and fitted regression lines by sale status for Sale 8990 (*-Sold Items, o-Unsold Items).

distribution with mean a and covariance matrix b . Let $S_i = \mathbf{I}(D_i \geq 0)$ be the response indicator, i.e., Y_i is observed when $S_i = 1$ and Y_i is missing when $S_i = 0$. The main interest is estimating the outcome equation

$$E(Y_i|\mathbf{X}_i) = \beta^T \mathbf{X}_i + \sigma \epsilon_i \quad \epsilon_i \stackrel{i.i.d.}{\sim} N(0, 1), \tag{3.2}$$

for the complete data.

This model plays a central role in the econometrics literature and it has been called Type II Tobit model (Amemiya (1984)) or stochastic censoring model (Little and Rubin (1987)). Various extensions and reparametrizations of the original models have been developed, of which the parametrization by Little and Rubin (1987) is useful here. The probability of Y_i being observed is

$$P(S_i = 1|\mathbf{X}_i, Y_i) = P(D_i \geq 0|\mathbf{X}_i, Y_i) = \Phi(\gamma^T \mathbf{X}_i + \delta Y_i), \tag{3.3}$$

where $\gamma = (\xi - \beta\rho\sigma^{-1})/\sqrt{1 - \rho^2}$ and $\delta = \rho/(\sigma\sqrt{1 - \rho^2})$. The selection equation (3.3) describes the missing-data mechanism explicitly. When $\delta = 0$, then $\rho = 0$, the missingness is ignorable conditional on the observed data. For ignorable missing data, the EM algorithm (Dempster, Laird and Rubin (1977)) and various extensions (Meng and Rubin (1991, 1993) and Meng and van Dyk (1997)) have been proposed for obtaining the maximum likelihood estimates. For the art auction data, the Wald tests of $H_0 : \delta = 0$ vs. $H_1 : \delta \neq 0$ in (3.7) are very significant (p -value < 0.0001) for all auction sales. This indicates a strong selection bias.

Selection models have been widely used in economics and the social sciences. The selection model with a logit selection equation has been considered by Greenlees, Reece and Zieshang (1982), Lee (1983), Olsen (1982), and others. Diggle and Kenward (1994) discussed models for informative drop-outs in longitudinal data analysis. There is a considerable literature on the nonparametric or semi-parametric estimation of selection models (Vella (1998); Das, Newey and Vella (2003)). Most of the estimation methods are extensions of Heckman's 1976; 1979 two-step approach based on the conditional expectation of Y ,

$$E(Y_i|X_i, S_i = 1) = \beta^T \mathbf{X}_i + \sigma W_i, \quad (3.4)$$

where $W_i = E(\epsilon_i|X_i, S_i = 1)$ is a correction term. For the normal selection model, $W_i = \phi(\alpha^T \mathbf{X}_i)/\Phi(\alpha^T \mathbf{X}_i)$, the inverse Mills' ratio. The two-step procedure uses the estimated correction term, \widehat{W}_i , as an extra regressor in (3.4). The actual implementation of the two-step procedure requires that the predictors in the outcome equation (3.2) and the selection equation (3.3) are not identical or have no collinearity (Vella (1998)). Consequently, the two-step procedure is *not* applicable in our situation, as the pre-sale estimate is the *only* predictor variable available to us and is used in both equations.

Let $X_i = \log(P_i)$, $Y_i = \log(A_i)$, $Z_i = \log(R_i)$, and let $S_i = I(A_i \geq R_i) = I(Y_i \geq Z_i)$ indicate whether the i th item is sold (1) or not (0). The pre-sale prediction P_i , hence X_i , is known priori to the auction sale. While the sale status S_i is always observed, Y_i is observed only when $S_i = 1$. The relationship between (X, Y, Z) for the art auction data can be modeled as:

$$Y_i = \beta_0 + \beta_1 X_i + \sigma \epsilon_{1i} \quad (3.5)$$

$$Z_i = \xi_0 + \xi_1 X_i + \sigma^* \epsilon_{2i}, \quad (3.6)$$

where $(\epsilon_{1i}, \epsilon_{2i}), i = 1, \dots, n$, are independent. Let $D_i = (Y_i - Z_i)/\sigma^*$. If $\epsilon_{2i} \sim N(0, 1)$, then the probability that an item sells is

$$P(S_i = 1|X_i, Y_i) = P(D_i \geq 0|X_i, Y_i) = \Phi(\gamma_0 + \gamma_1 X_i + \delta Y_i), \quad (3.7)$$

where $\gamma_0 = -\xi_0/\sigma^*$, $\gamma_1 = -\xi_1/\sigma^*$, and $\delta = 1/\sigma^*$.

Let $\mathbf{X} = (X_i; i = 1, \dots, n)$ and $\mathbf{S} = (S_i; i = 1, \dots, n)$. For convenience, we record the data so that the first n_0 items are not sold and $\mathbf{Y}_{\text{obs}} = (Y_i; i = n_0 + 1, \dots, n)$ are the observed sale price (final bids) for the last $n - n_0$ sold items. If $\theta = (\beta_0, \beta_1, \sigma)$ and $\psi = (\gamma_0, \gamma_1, \delta)$, then the likelihood function for the released auction data $(\mathbf{Y}_{\text{obs}}, \mathbf{S}, \mathbf{X})$ is

$$L(\theta, \psi|\mathbf{Y}_{\text{obs}}, \mathbf{S}, \mathbf{X}) = \prod_{i=1}^{n_0} P(S_i = 0|X_i) \prod_{i=n_0+1}^n f(Y_i, S_i = 1|X_i), \quad (3.8)$$

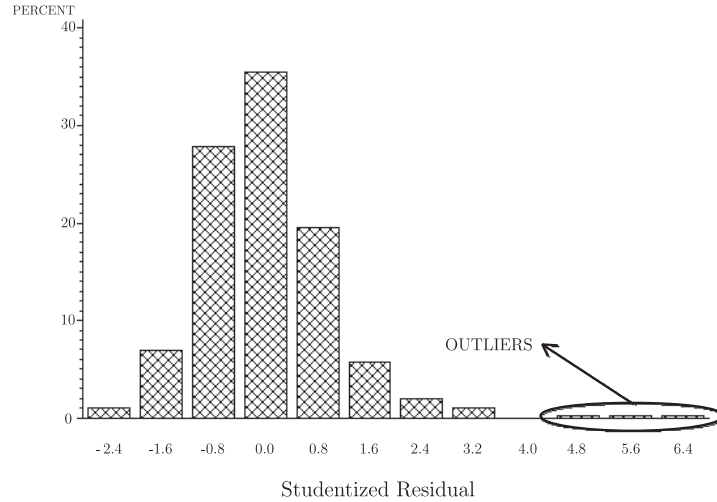


Figure 3.2. The discrete histogram of the regression residuals for Sale 9038.

where $f(Y_i, S_i = 1|X_i) = P(S_i = 1|X_i, Y_i)f(Y_i|X_i)$ and $P(S_i = 0|X_i) = \int P(S_i = 0|X_i, y)f(y|X_i)dy$. In the normal selection model, the marginal distribution $f(Y_i|X_i)$ is normal, i.e., the error term $\epsilon_{1i} \sim N(0, 1)$. The maximum likelihood estimates (MLE) can be obtained by the Newton-Raphson method (see Appendix 1). The computation can be implemented using numerical analysis software MAPLE V (Release 4) or SAS PROC NLP for optimization of a continuous non-linear function. These programs calculate the required first and second derivative symbolically. In the data analysis, the starting values are the estimates from the regression for sold items as starting values, and the stopping rule is either one reaches the maximum number of iterations 100 or the maximum difference of the estimates between consecutive iterations is below 10^{-4} .

When the normal selection model is used to analyze the art auction data, we find that the residuals $\epsilon_{1i}, k = 1, \dots, n$, in the response equation have several outliers, as some items received very high bids. Figure 4.3 presents the distribution of these residuals for auction sale 9038. The skewness measures of the regression residuals for the five auction sales are (0.6, 1.4, 0.7, 0.8, 1.0), respectively. After removing the upper 5% of the residuals, the residual distributions are almost symmetric, with skewnesses (0.1, 0.2, 0.0, -0.1, 0.1). In order for the parameters θ and ψ to be identifiable, some parametric assumptions concerning the distributions of the errors $(\epsilon_{1i}, \epsilon_{2i})$ need to be made. Estimates of δ in the selection equation and θ in the outcome equation are affected by the model assumptions as well as the observed data. Furthermore, estimates from the normal selection models are not robust to misspecification of the error distribution (Little (1985)). For example, Glynn, Laird and Rubin (1986) allowed the residuals to follow a t

or mixture normal distribution, and showed that the estimates from the normal selection model were seriously biased. Copas and Li (1997) used the parameter δ to reflect the degree of non-randomness of the sample and assessed the sensitivity of inferences to small non-zero values of δ .

The t -distributions provide a useful extension of the normal for statistical modeling of data involving errors having heavier tails. Gastwirth (1970) and Lange, Little and Taylor (1989) discuss robust methods using the t distribution. Thus, we replace the normal distribution by the t -distribution for the error term in (3.2) in our analysis. The t density is

$$f_t(y|\mu, \sigma^2, \nu) = \frac{\Gamma((\nu + 1)/2)}{\Gamma(1/2)\Gamma(\nu/2)\nu^{1/2}\sigma} \times \left(1 + \frac{(y - \mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2},$$

and the likelihood function is obtained by replacing $f(Y_i|X_i; \theta)$ with $f_t(Y_i|\beta_0 + \beta_1 X_i, \sigma^2, \nu)$ and approximating $P(S_i = 0|X_i)$ by Gauss-Laguerre integration (see Appendix 2). The t -distributions with lower degrees of freedom have higher resistance to outliers, but may have lower efficiency. Lange et al. (1989) fixed the degrees of freedom at a pre-determined value (such as 4). However, art auction data often contains extremely high bids for a few sold items, and in the analysis we found that the t_2 distribution worked well and so present the results from the t_2 selection model. One could include the degrees of freedom as a parameter to be estimated. But the calculations would become more complex.

4. Evaluating the Utility of the Selection Models

In this section, we evaluate the fit of the selection models by comparing the estimated and actual final bids for the unsold items. Then we compare the estimates of forecast error measures from different methods using only the publicly released auction data and their values calculated from the complete data.

4.1. Evaluating the estimated final bids for the unsold items

To evaluate the fit of the selection models, we focus on the unsold items. For each unsold item, we compare its imputed final bid conditional on the sale status, $A_i^{(0)}$, to the actual highest bid $A_i, i = 1, \dots, n_0$, which is usually not published by the auction house. The final bid for the i -th unsold item is estimated by

$$A_i^{(0)} = E(A_i|S_i = 0) = \frac{E\{A_i I(S_i = 0)\}}{P(S_i = 0)}. \quad (4.1)$$

The details of the calculation are in the supplementary document.

Table 4.4 reports several summary measures of the difference. The first measure is the ratio of the imputed total to the actual total of the bids for the unsold items. The second and third measures are the average bias and

Table 4.4. Comparison of the imputed final bids for the unsold items obtained from different selection models.

| Auction | Selection Model | $\frac{\sum A_i^{(0)}}{\sum A_i}$ | $\frac{\sum(A_i - A_i^{(0)})}{n_0}$ | $\frac{\sum A_i - A_i^{(0)} }{n_0}$ | $\frac{1}{n_0} \sum \left \frac{A_i - A_i^{(0)}}{A_i} \right $ |
|---------|-----------------|-----------------------------------|-------------------------------------|--------------------------------------|---|
| 3,850 | Normal | 0.829 | -125.21 | 154.87 | 0.197 |
| | t_2 | 1.019 | 13.60 | 93.77 | 0.138 |
| 6,371 | Normal | 0.674 | -1612.48 | 1789.83 | 0.277 |
| | t_2 | 0.904 | -473.69 | 1318.58 | 0.190 |
| 8,990 | Normal | 0.818 | -1173.86 | 1300.95 | 0.202 |
| | t_2 | 1.056 | 358.52 | 789.84 | 0.146 |
| 9,028 | Normal | 0.806 | -3221.45 | 3337.48 | 0.212 |
| | t_2 | 1.046 | 758.90 | 1678.41 | 0.159 |
| 9,038 | Normal | 0.828 | -1121.65 | 1250.77 | 0.217 |
| | t_2 | 1.099 | 648.36 | 1035.78 | 0.136 |

the average absolute error. The last measure is the average relative imputation error. From the third column of Table 4.4, we see that the normal selection model substantially underestimates the total of all the actual bids received for the unsold items. The imputed total derived from the normal selection model ranges from 67% to 83% of the actual total final bids for the unsold items. The selection model based on the t_2 distribution performed better for every auction. It yields a total of the imputed prices ranging from 90% to 110% of the total actual bids. In contrast the estimates of the highest bids for the unsold items obtained from the normal selection model were usually below their actual value. Both measures of bias in columns 4 and 5 in Table 4.4 are noticeably lower for those from t_2 error model and the relative imputation errors for unsold items in the last column are much lower.

4.2. Evaluating the forecast error measures after using the selection model

The forecast error measures *AREA* and *AREP* are estimated using different selection models. The estimates from an appropriate selection model applied to the publicly released data for the objects that sold should be close to those values. First, the selection models are applied to the publicly released data. Then *AREA* and *AREP* are calculated after including the imputed final bids for the unsold items obtained from the selection models, $A_i^{(0)}$. For comparison, the estimates of *AREA* and *AREP* without adjustment for selection bias from the full data as well as publicly released data are also given in Table 4.5. For example, the estimates of *AREA* and *AREP* for full data are shown on the “All” lines.

Table 4.5 demonstrates that the measures of forecast error are well approximated by the t_2 selection model. Both estimated measures from this model are

Table 4.5. Evaluation of forecast errors using different models for the publicly released data.

| Auction | Items used | Adjustment | <i>AREA</i> | <i>AREP</i> |
|---------|------------|------------------------|-------------|-------------|
| 3,850 | All | No adjustment | 0.325 | 0.295 |
| | Sold | No adjustment | 0.224 | 0.277 |
| | Sold | Normal selection model | 0.416 | 0.333 |
| | Sold | t_2 selection model | 0.296 | 0.287 |
| 6,371 | All | No adjustment | 0.577 | 0.450 |
| | Sold | No adjustment | 0.339 | 0.426 |
| | Sold | Normal selection model | 0.813 | 0.497 |
| | Sold | t_2 selection model | 0.545 | 0.438 |
| 8,990 | All | No adjustment | 0.353 | 0.327 |
| | Sold | No adjustment | 0.290 | 0.315 |
| | Sold | Normal selection model | 0.407 | 0.346 |
| | Sold | t_2 selection model | 0.333 | 0.323 |
| 9,028 | All | No adjustment | 0.395 | 0.307 |
| | Sold | No adjustment | 0.288 | 0.276 |
| | Sold | Normal selection model | 0.475 | 0.333 |
| | Sold | t_2 selection model | 0.360 | 0.299 |
| 9,038 | All | No adjustment | 0.329 | 0.367 |
| | Sold | No adjustment | 0.280 | 0.364 |
| | Sold | Normal selection model | 0.386 | 0.384 |
| | Sold | t_2 selection model | 0.312 | 0.360 |

close to their values calculated from the complete data. The normal selection model overestimates the measures of forecast error, which suggest the pre-sale estimates are less accurate than they truly are. This is expected from the finding that the estimated highest bids, obtained from the normal selection model, were too low. For each auction, the second line in the table reminds us that *AREA* and *AREP* calculated from the publicly released data are underestimated.

4.3. Checking the model adequacy by residual plot

The goodness-of-fit of a model is commonly assessed by the plots of residuals. Table 4.4 presents several accuracy statistics for the residual differences between the imputed ($\hat{A}^{(0)}$) and the observed final bids (A) for the *unsold* items. Figure 4.3 presents the plots of the residuals $\log(\hat{A}^{(0)}) - \log(A)$ vs. $\log(A)$ from the selection models for the *unsold* items for Sale 8,990. The residual plots for other auction sales follow similar pattern. The dots are the residuals from the normal selection model and the circles are the residuals from the t -selection model. Cubic spline curves for the residuals are also shown in Figure 4.3, where the solid (dash) line is based on residuals from the normal (t) selection model. From the plots in Figure 4.3 and the fourth column in Table 4.4, it is clear that the imputed

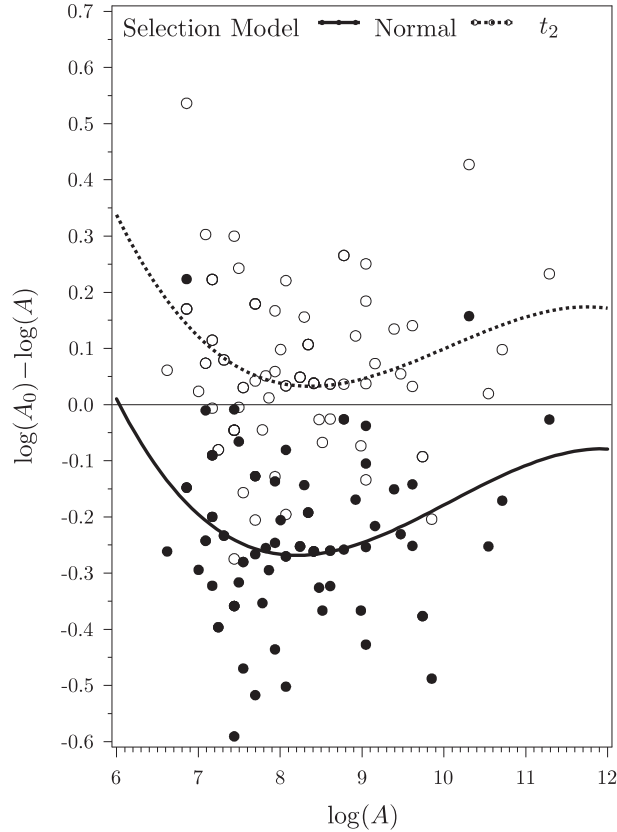


Figure 4.3. Plot of the residuals between observed and imputed final bids for *unsold* items for Sale 8990 (\bullet : normal selection model, \circ : t -selection model).

final bids based on normal selection models underestimate the actual final bids for unsold items, while the residuals for the t -selection model are closer to 0.

5. Simulation

To examine the robustness of the normal selection model and the t -selection model, we performed a small simulation study. The primary equation of interest is

$$Y = \beta_0 + \beta_1 X + \sigma \epsilon.$$

In art auctions, $Y = \log(A)$ is the highest bid and $X = \log(P)$ is the pre-sale prediction. Based on the estimated regression coefficients in Table 2.3, we set $\beta_1 = 1$ and $\sigma = 0.5$. We assume that the predictor x is standardized and generated from a standard normal distribution. To introduce contamination in the response y , we assumed that $\beta_0 = b_0$ with probability $1 - c$ and $\beta_0 = 3$ with contamination probability c , where b_0 takes three possible values 0, 0.2

Table 5.6. Biases of the parameters β_0 and β_1 for normal and t_2 selection models.

| n | β_0 | R/P | No contamination | | | | Contamination rate $c = 0.1$ | | | | |
|-----|-----------|-------|------------------|-----------|-----------|-----------|------------------------------|-----------|-----------|-----------|-------|
| | | | Normal | | t_2 | | Normal | | t_2 | | |
| | | | β_0 | β_1 | β_0 | β_1 | β_0 | β_1 | β_0 | β_1 | |
| 200 | 0 | 0.4 | 0.044 | 0.020 | 0.023 | 0.026 | 0.352 | 0.021 | 0.060 | 0.025 | |
| | | 0.6 | 0.142 | 0.016 | 0.103 | 0.021 | 0.480 | 0.016 | 0.144 | 0.020 | |
| | 0.2 | 0.4 | 0.021 | 0.021 | 0.009 | 0.027 | 0.324 | 0.023 | 0.045 | 0.026 | |
| | | 0.6 | 0.083 | 0.018 | 0.053 | 0.024 | 0.402 | 0.019 | 0.091 | 0.023 | |
| | 0.4 | 0.4 | 0.009 | 0.021 | 0.003 | 0.027 | 0.311 | 0.024 | 0.039 | 0.027 | |
| | | 0.6 | 0.045 | 0.020 | 0.024 | 0.026 | 0.353 | 0.021 | 0.061 | 0.025 | |
| | 500 | 0 | 0.4 | 0.044 | 0.013 | 0.024 | 0.016 | 0.350 | 0.015 | 0.060 | 0.016 |
| | | | 0.6 | 0.143 | 0.010 | 0.103 | 0.012 | 0.478 | 0.011 | 0.143 | 0.012 |
| 0.2 | | 0.4 | 0.021 | 0.013 | 0.009 | 0.016 | 0.323 | 0.016 | 0.045 | 0.016 | |
| | | 0.6 | 0.084 | 0.011 | 0.053 | 0.014 | 0.400 | 0.014 | 0.091 | 0.015 | |
| 0.4 | | 0.4 | 0.009 | 0.013 | 0.003 | 0.017 | 0.309 | 0.016 | 0.039 | 0.017 | |
| | | 0.6 | 0.045 | 0.013 | 0.024 | 0.016 | 0.351 | 0.015 | 0.060 | 0.016 | |

and 0.4. When $c = 0$, there is no contamination and when $c = 0.1$, 10% of the responses are contaminated with extremely high values. To mimic the art auction data, we assumed that the probability of Y being observed is $P(S = 1|X, Y) = \Pr(Y \geq Z) = \Phi(Y - Z)$, where $Z = \log(R)$ is the logarithm of the reserve price. We considered two possible rules for setting the reserve prices: $R/P = 0.6$ or $R/P = 0.4$. Two sample sizes $n = 200$ and 500 were used in the simulation, and the missing rate ranged from 11% to 31%.

We then fit the normal selection model and the selection model with t_2 distributions to the simulated data and calculated the biases of the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ in the primary outcome equation. The average biases of $\hat{\beta}_0$ and $\hat{\beta}_1$ with 1,000 simulations are summarized in Table 5.6.

When there was no contamination, the biases of β_0 and β_1 for both selection models were very small and similar, although the bias for $\hat{\beta}_1$ was slightly greater for the t_2 model. When there was 10% contamination, i.e., about 10% of the items received extremely high bids, the biases for β_1 from both selection models were still small and quite close. However, the biases for β_0 from the normal selection model were noticeably higher than those from the t_2 selection model. Overall, the t -selection models showed much more resistance to the outliers than the normal selection models.

6. Discussion

The Heckman selection model is widely used, often without questioning its validity, in econometrics. Using the art auction data, we showed that this model

may fail in the presence of outliers and that the selection model with a t_2 distribution is a useful alternative. In general, the t -distribution with 2 degrees of freedom (DOF) may not be optimal and the DOF could also be estimated from the data.

Measures of the forecast error of art auction pre-sale predictions, relying solely on the published data for sold items, have a noticeable selection bias. As the final bids for the unsold items are substantially lower than the pre-sale predictions, without correcting for this, one underestimates the forecast error of the pre-sale prices. Thus, participants in the art market should allow for this added uncertainty when using the predicted price interval to determine their bids.

The more important *AREA* measure comparing the pre-sale estimates to the actual prices is more severely affected than the *AREP* measure. It is also used in the evaluation of company earning forecasts (Jaggi (1978)). Gastwirth (1979) showed that the *AREA* is equivalent to the coefficient of dispersion (CD) used to assess the forecast error and fairness of real estate tax assessments. The *AREAs* of the art auction are in the same range as those of the real estate assessments.

In auction sales, the pre-sale prediction and reserve price are determined by art experts. The first should be a good-faith estimate and the second provides a "reasonable" lower limit for the sale price. There is no upper limit to the bidding. Several studies showed that the distributions of the final highest bids are right-skewed (McAndrew and Thompson (2007)), so appropriate transformation or skewed distributions may be applied to the data. There are a few outliers in the art auction data, which may be due to unusual competition (Lance, Klein and Weiss (1987)). The sold items with extremely high bids produce large residuals, yielding a large standard residual variance. The selection model that assumes normally distributed random errors is highly influenced by the extremely high bids for some sold items; as a result, it imputes overly negative residuals to the missing (unsold) items. The low imputed bids for the unsold items will overestimate the difference between pre-sale prediction and the final bids for the unsold items. Kenward (1998) observed a similar phenomena in the analysis of data on mastitis in dairy cows (see his Figure 1).

Since both the outcome and selection equations in the art auction context have the same predictor, the pre-sale estimate provided by the auction house, semi-parametric methods are not applicable (Vella (1998)). Replacing the normal distribution in the classical model of Heckman by the t_2 distribution, which is known to be more resistant to outliers than the normal (Lange et al. (1989) and Pinheiro, Liu and Wu (2001)), allows for the larger variability inherent in auction data. Unlike the usual applications of selection models where one does not have data on the items not "selected" for the sample, the highest bids obtained for the

unsold items were available and were used to verify that the selection model with errors following a t_2 distribution yielded much better estimates of the missing values and the measures of prediction accuracy. It is noteworthy that these improved results held for auctions of several types of art, i.e. Chinese painting, 20th Century prints and jewelry.

Recently, the skew t -distribution (Azzalini and Capitanio (2003)) has been shown to be effective in modeling the presence of skewness effects and possible heavy tails simultaneously. Arellano-Valle, Branco and Genton (2006) showed that the skewed distributions may arise from selection. For the extension of the approach developed here, one may consider the skewed t distribution for the final bids of the sold items because of the selection effect. The analysis is more computationally intensive as both the degrees of freedom and skewness should be estimated.

Acknowledgement

This research was supported in part by NSF grant SES-0317956 awarded to the George Washington University. The authors wish to thank Mr. John Cavaliero of Cavaliero Fine Arts, an art dealer in New York, for his interest in this project and for introducing us to the experts in Christie's, and Mr. Chris Eykyn and Ms. Loren Rubin of Christie's for kindly providing us the three print catalogs with the complete data.

References

- Amemiya, T. (1984). Tobit model: a survey. *J. Econometrics* **24**, 3-61.
- Arellano-Valle, R., Branco, M. D. and Genton, M. G. (2006). A Unified view on skewed distributions arising from selections. *Canad. J. Statist.* **34**, 581-601.
- Azzalini, A. and Capitanio, A. (2003). Distribution generated by perturbation of symmetry with emphasis on a multivariate skewed t distribution. *J. Roy. Statist. Soc. Ser. B* **65**, 367-389.
- Copas, J. B. and Li, H. G. (1997). Inference for non-random samples. *J. Roay Statist. Soc. Ser. B* **59**, 55-95.
- Das, M., Newey, W. and Vella, F. (2003). Nonparametric estimation of sample selection models. *Rev. Econ. Studies* **70**, 33-58.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy Statist. Soc. Ser. B* **39**, 1-37.
- Diggle, P. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *App. Statist.* **43**, 49-93.
- Gastwirth, J. L. (1970). On robust rank tests, *Proc. 1st International Symp. on Nonparam. Techniques in Statist. Inf.* (Edited by by M. L. Puri). Cambridge University Press.
- Gastwirth, J. L. (1979). Evaluating the accuracy of art auction estimates. *Proc. Business and Econ. Statist. Sec.* American Statistical Association, 150-153.

- Glynn, R. J., Laird, N. M. and Rubin, D. B. (1986). Mixture modeling versus selection modeling with nonignorable nonresponse. In *Drawing Inferences from Self-selected Samples* (Edited by H. Wainer), 115-142. Springer, New York.
- Greenlees, W. S., Reece, J. S. and Zieshang, K. D. (1982). Imputation of missing values when the probability of nonresponse depends on the variable being imputed. *J. Amer. Statist. Assoc.* **77**, 251-256.
- Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Ann. Econ. and Social Measurement* **5**, 475-492.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153-160.
- Inde, V. R. (1998). *Art in the Courtroom*. Praeger, Westport, CT.
- Jaggi, B. (1978). Comparative accuracy of managements annual earning's forecasts. *Finan. Management* **7**, 24-32.
- Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statist. Medicine* **17**, 2723-2732.
- Lance, B., Klein, J. D. and Weiss, L. W. (1987). The price effects of increased competition in auction markets. *Rev. Econom. Statist.* **69**, 24-32.
- Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution *J. Amer. Statist. Assoc.* **84**, 881-896.
- Lee, L. F. (1983). Generalized econometric models with selectivity. *Econometrica* **51**, 507-512.
- Little, R. J. A. (1985). A note about models for selectivity bias. *Econometrica* **53**, 1469-1474.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley, New York.
- McAndrew, C. and Thompson, R. (2007). The collateral value of fine art. *J. Banking and Finan.* **31**, 589-607.
- Meng, X-L. and Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J. Amer. Statist. Assoc.* **86**, 899-909.
- Meng, X-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm. *Biometrika* **80**, 267-278.
- Meng, X-L. and van Dyk, D. A. (1997). The EM algorithm-an old folk song sung to a fast new tune (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 511-567.
- Olsen, R. J. (1982). Distributional tests for selectivity bias and a more robust likelihood estimator. *Internat. Econ. Rev.* **24**, 181-196.
- Pinheiro, J., Liu, C. and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t -distribution. *J. Comput. Graph. Statist.* **10**, 249-276.
- Vella, F. (1998). Estimating models with sample selection bias: a survey. *J. Human Resources* **33**, 1-167.

Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, MD 20904, U.S.A.

E-mail: yubi@mail.nih.gov

Department of Statistics, The George Washington University, Washington DC 20052, U.S.A.

E-mail: jlgast@gwu.edu

(Received July 2007; accepted December 2008)