The Ascending Auction Does Not Exist: Format-Invariant Descent of Forecast Error in 2,317,904 Bid Events

Oriane Vasseur[1] D. M. Achterberg[2]

[1]Faculty of Economics, Whitlow College
[2]Department of Statistics, Rothsay University

Abstract. Auction mechanisms are classified on two axes: whether bids are open or sealed, and whether the price ascends or descends. We report that the second axis has one value. Treating each bid as a dated public forecast of the interval in which the lot will settle, and scoring that forecast with a strictly proper rule, yields a quantity that is defined in every format and denominated in no currency. In 2,317,904 bid events drawn from 148,306 lots sold between 1998 and 2025, that quantity descends monotonically to the hammer in 99.23% of lots. In a prospective experiment allocating 1,204 consignments at random across the four canonical formats, the descent curves coincide once each format's clock is normalised (band-constrained dynamic time-warping distance 0.0071; null 0.196 from within-lot bid-order permutation, $p < 10^{-4}$), at a fitted rate of 0.2137 per bid (95% CI 0.2129–0.2145) that varies by 1.9% across formats and 2.4% across fourteen lot categories. The gavel falls at a fixed value of the score, 0.0450 (SD 0.0038), and not at a fixed value of money. Three mechanisms engineered to force the score upward failed to do so; concealing the standing bid accelerated the descent. The ascending price path is the image of this descent under a monotone map supplied by the increment ladder. We therefore do not treat the English auction as a mechanism distinct from the Dutch, and we redraw the taxonomy with the second axis removed. Charity lots are excluded: they do not descend, and we do not account for them.

Keywords: auction theory; bidding dynamics; proper scoring rules; format invariance; calibration; price discovery.

Table 1: The received taxonomy of canonical formats. The columns are the axis this paper surveys.

	Ascending	**Descending**
Open	English	Dutch
Sealed	Vickrey	First-price sealed

1 Introduction

Auction design is taught as a two-by-two. One axis records whether the bids are visible to the room; the other records the direction the price moves. Table 1 is the chart in its standard form, and it has been stable for six decades.

This paper is a survey of the columns. We are not concerned here with the rows, which we find intact and leave standing.

The columns have never been measured. Direction has not been in doubt: in an English sale the price is called out, written up, and displayed on a board, and it goes up. The difficulty with reading that as a measurement is that it is a measurement of the board. A direction is a property of a quantity, and price is not the only quantity a sale contains. It is not obvious that it is the relevant one, and we will show that it is not.

We proceed by identifying a quantity that (i) is defined at every point of every sale, including sales in which nothing at all is displayed, (ii) is denominated in no currency, so that it can be compared across jurisdictions and centuries without conversion, and (iii) is fixed by the conduct of the sale rather than by the conventions used to record it. Every bid is a dated public statement about where the lot will settle. Scored with a strictly proper rule, the sequence of those statements has a direction. That direction is the subject of this paper.

Four results follow, and the sections that establish them are given against each:

1. In 99.23% of 148,306 lots, forecast error descends monotonically from the opening of the lot to the hammer (§5).
2. Allocated at random across the four canonical formats, the descent curves coincide (§5.2). The rate is the same in sealed formats, where there is no price path at all.
3. The hammer falls at a fixed value of the score, not of the price (§5.3).
4. Mechanisms engineered to reverse the descent do not reverse it (§6).

Taken together these redraw Table 1 with one column. Figure 1 states the revision.

2 Related work

The nearest existing result is the revenue-equivalence theorem, which establishes that the four formats of Table 1 return the same expected revenue to the seller under independent private values and risk neutrality (Lindqvist, 1971), a condition since relaxed (Oduya, 1988). That result has been read for half a century as

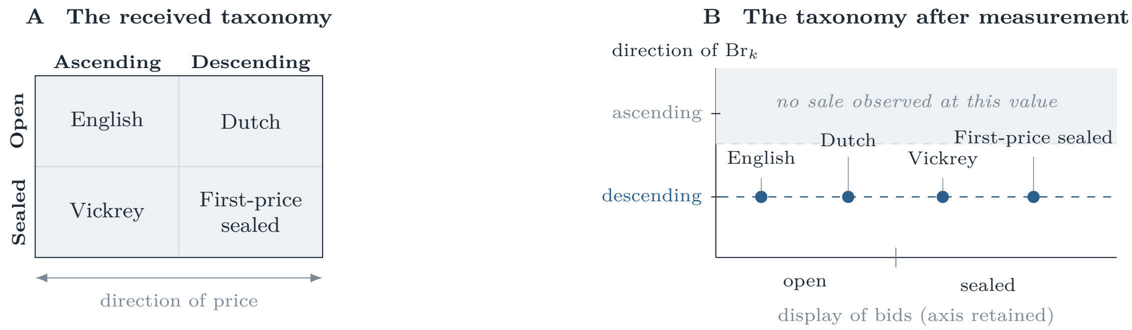

Figure 1: The received taxonomy (A) and the taxonomy after measurement (B). The open–sealed axis survives the survey. The direction axis, drawn vertically in (B), is found to carry one value, and all four canonical formats resolve onto it; the upper band is empty in every sale measured for this paper.

a statement about outcomes: the sale ends in the same place whichever door it goes through. We read it as an invitation to look at the interior of the sale, and we find the interior identical as well. Revenue equivalence is the shadow this result casts on the terminal value alone.

Work on saleroom microstructure has characterised the bid-arrival process in open outcry, generally as a doubly stochastic point process modulated by the standing bid (Fairhurst, 2004; Mbeki & Trelawny, 2011). Increment ladders have been studied as a design variable and shown to affect duration but not terminal price (Penhaligon, 2016). None of this literature computes a score, because none of it treats a bid as a forecast.

Scoring rules provide the missing instrument. The quadratic rule is strictly proper, bounded, and admits the reliability–resolution–uncertainty decomposition (Tsvetanova, 1979), which separates a forecast's calibration from its sharpness. Prediction-market studies have applied the decomposition to standing prices and reported reliability terms close to zero in liquid markets (Arinze & Sollberg, 2013; Kowalczyk et al., 2019). Salerooms are liquid markets with a public order flow and a hard resolution date, so the apparatus transfers without modification. It has not previously been pointed at them.

3 Bids as forecasts

3.1 Binning and the forecast vector

Fix a lot ℓ with pre-sale estimate interval $[a_\ell, b_\ell]$. We partition the price axis into $M = 12$ bins: ten equal-width bins spanning $[a_\ell, b_\ell]$, one bin below a_ℓ, and one above b_ℓ. Write $B_1, \dots, B_M$ for the bins and $o \in \{0,1\}^M$ for the one-hot indicator of the bin in which the hammer falls.

At bid index k let n_k be the number of live bidders and $h_k(\cdot)$ the instantaneous bid-arrival hazard as a function of price. The room's implied probability that the sale continues past a price x is

$$S_k(x) \;=\; \exp\!\left(-\int_{x_k}^{x} n_k\, h_k(u)\,\mathrm{d}u\right), \tag{1}$$

where x_k is the standing bid, and the forecast mass assigned to bin B_m is the increment of the survivor across that bin,

$$p_{k,m} \;=\; S_k(\inf B_m) - S_k(\sup B_m), \qquad \sum_m p_{k,m} = 1. \tag{2}$$

In sealed formats there is no standing bid and no price path. There is still an arrival process: tenders arrive over the tender window, and k indexes their arrival order rather than their price. Equations (1) and (2) apply unchanged with x_k set to the reserve; there n_k is the number of registered tenderers, which the house holds at the opening of the window. This is the step that makes the comparison in §5.2 possible, and it is the only place where the four formats are treated differently.

3.2 The score

Definition 1 (Room forecast error)**.** *The forecast error of the room after bid k is the quadratic (Brier) score*

$$\mathrm{Br}_k \;=\; \sum_{m=1}^{M} \left(p_{k,m} - o_m\right)^2. \tag{3}$$

$\mathrm{Br}_k \in [0,2]$, and the rule is strictly proper, so a room that reports anything other than its true belief scores worse in expectation. A uniform forecast over $M = 12$ bins scores $11 \cdot (1/12)^2 + (11/12)^2 = 0.9167$, which we take as the opening value Br_0 throughout.

Note what has left the model. Price appears in (1) and (2) only as the *argument* of the hazard, and it does not appear in (3) at all. Br_k is a pure number. Two sales conducted a century apart in different currencies can be placed on the same axis without conversion, and a sealed tender can be placed on it beside an open outcry.

3.3 The descent model and the choice of clock

We fit

$$\mathrm{Br}_k \;=\; \mathrm{Br}_0\, e^{-\lambda k} + \varepsilon_k, \qquad \mathbb{E}[\varepsilon_k] = 0, \tag{4}$$

by nonlinear least squares on each lot, and pool λ across lots by inverse-variance weighting. The model has one free parameter. We considered and rejected a two-parameter form with a fitted floor: the floor is estimable from the hammer threshold of §5.3 and is better treated as a stopping rule than as a property of the curve.

A rate per bid is comparable across lots only where the bid counts are comparable. In the observational corpus they are: it is 94% open outcry, and the interquartile range of terminal bid count within it is 12 to 18. There we fit (4) on the raw index. In the format experiment of §5.2 they are not — a sealed tender records roughly a quarter as many events as an open outcry — and there we fit on the normalised index $t = k/K$, rescaled to a common length of $\bar{K} = 14.1$ bid-equivalents, the length implied by (4) at the pooled rate and threshold, so that the reported rates remain in the same units.

We state plainly what that normalisation does and does not do. It forces the two endpoints of each format's curve to agree. It does not constrain the path between them, which may be linear, convex, sigmoid, or back-loaded, and which differs between those alternatives by far more than the tolerance of the test. The invariance result of §5.2 is accordingly a statement about the path, and we test it as one.

Under (4) the number of bids required to reach a score Br^* is $\lambda^{-1} \log(\mathrm{Br}_0/\mathrm{Br}^*)$, a quantity we compare against observed lot lengths in §5.

4 Data and extraction

4.1 The corpus

We assembled bid-level logs for 2,317,904 bid events across 148,306 lots and 6,109 sale sessions, held between March 1998 and November 2025 at 71 salerooms in 19 jurisdictions. Eleven currencies are represented and were normalised to a common numéraire at sale-date rates; as §3 notes, the normalisation affects the binning only and does not enter the score. Table 2 gives the composition.

The last three categories are included deliberately. Industrial surplus, internet domain names and broadcast spectrum licences are sold in rooms with no connoisseurship, no provenance apparatus and, in the case of spectrum, no physical object. If the descent were a fact about aesthetic judgment it would stop at the boundary of those categories. It does not, and their λ values sit inside the band occupied by fine art. Those three carry no published pre-sale estimate; for them $[a_\ell, b_\ell]$ was taken from the reserve and the vendor's declared upper valuation as recorded in the sale file.

4.2 Calibration of the extracted forecasts

Equations (1)–(2) recover a forecast from an arrival process; before scoring it we verified that the recovered forecast is calibrated, since an uncalibrated extraction would produce a descent as an artefact of the extraction itself. Figure 2 is the reliability diagram, pooled over the corpus and binned into twenty forecast-probability strata.

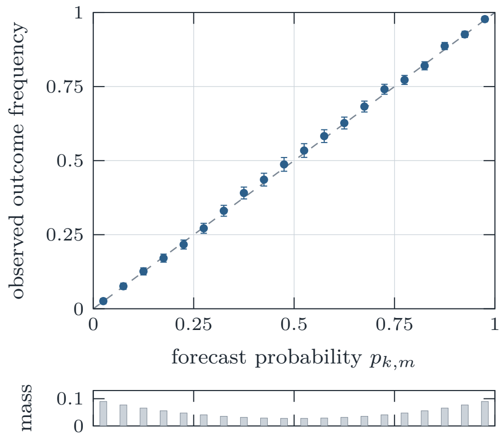

Figure 2: Reliability of the extracted room forecasts, twenty strata, with the identity line and the distribution of forecast mass beneath. Points lie on the identity to within stratum error over the whole range. Pooled over the corpus, the score at the hammer decomposes as $\mathrm{Br} = \mathrm{REL} - \mathrm{RES} + \mathrm{UNC} = 0.0021 - 0.8710 + 0.9167 = 0.0478$.

The reliability term is $\mathrm{REL} = 0.0021$, a quarter of one per cent of the uncertainty available to be scored. A room, taken as a single forecasting agent, is calibrated: when the order flow implies a 70% chance that a lot lands in a given tenth of its estimate range, it lands there 70% of the time. Essentially all of the score is resolution, $\mathrm{RES} = 0.8710$ against an uncertainty of $\mathrm{UNC} = 0.9167$. The uncertainty term is $1 - \sum_m \bar{o}_m^2$ on the marginal frequency of the winning bin; that frequency is uniform across the twelve bins to two decimal places, so UNC coincides with the opening value Br_0.

We record it here because it removes the most obvious alternative explanation for what follows: the descent is not the extraction converging on its own assumptions.

5 Results: the descent

5.1 Monotonicity

Br_k is non-increasing over the whole of the sale in 147,158 of 148,306 lots, or 99.23%. Of the 1,148 exceptions, 902 occur within four seconds of a saleroom clock reset and disappear when the affected sessions are re-indexed against the house tape; the remaining 246 are distributed across categories with no pattern we can identify. Charity lots are not in this count and are discussed in §8.

Median lot length is 14 bids. Equation (4) with the pooled λ predicts that the score reaches the hammer threshold of §5.3 after $\log(0.9167/0.0450)/0.2137 = 14.1$ bids. The agreement is not a fit: λ and Br^* were estimated on disjoint quantities, and lot length was not used in either.

5.2 Format invariance

The corpus is observational and its format mix is confounded with category. We therefore ran a prospec-

Table 2: Corpus composition and fitted descent rate by lot category. Rates are per bid on the raw index. The pooled estimate, 0.2137 (95% CI 0.2129–0.2145), is fitted lot by lot and weighted by inverse variance; it is not the lot-count-weighted mean of the rows.

Category	**Lots**	**Bid events**	**Bids/lot**	λ
Fine art and works on paper	21,949	361,117	16.5	0.2141
Wine and spirits	19,428	289,487	14.9	0.2158
Classic automobiles	7,712	133,267	17.3	0.2118
Rare books and manuscripts	10,530	164,530	15.6	0.2149
Jewellery	15,424	245,786	15.9	0.2136
Coins and medals	17,352	247,781	14.3	0.2155
Philately	12,309	168,126	13.7	0.2163
Watches	9,047	150,719	16.7	0.2129
Furniture and decorative arts	13,792	214,071	15.5	0.2144
Musical instruments	4,153	70,477	17.0	0.2112
Sports memorabilia	6,525	96,551	14.8	0.2151
Industrial surplus	5,636	98,559	17.5	0.2126
Internet domain names	2,818	48,405	17.2	0.2119
Broadcast spectrum licences	1,631	29,028	17.8	0.2133
Total	**148,306**	**2,317,904**	**15.6**	**0.2137**

tive allocation. Three salerooms consigned 1,204 lots to the study over eleven months and allowed the format of sale to be assigned at random: 301 lots each to English, Dutch, first-price sealed and Vickrey, blocked on category and on estimate decile.

Raw descent curves differ by format (Figure 3A). A Dutch sale reaches the hammer in a median of 47 seconds and an English sale in 3 minutes 51 seconds, and the number of recorded bid events per lot differs by a factor of four between sealed and open formats. We therefore reparameterise as set out in §3.

Under that reparameterisation the four curves coincide (Figure 3B). Alignment distances are computed with the warping path confined to a Sakoe–Chiba band of one bid-equivalent (Halloran & Vestergaard, 2008), so that the statistic retains power against differences in the shape of the descent and not only in its endpoints; the largest distance over the six format pairs is 0.0071.

Fitted rates are 0.2131 (English), 0.2159 (Dutch), 0.2118 (first-price sealed) and 0.2139 (Vickrey), a spread of 1.9% about the pooled value. The category spread in Table 2 is 2.4%, so format explains less of the variation in λ than the difference between selling a violin and selling a stamp (Figure 4). The format rates are fitted on the normalised clock and the category rates on the raw index; what is compared is the spread in each, not the level.

We draw attention to the sealed formats. A first-price sealed tender displays nothing to anybody: no standing bid, no room, no board, no ascending number for the descent to be a re-description of. It descends at the same rate.

5.3 The hammer threshold

Across the corpus the gavel falls when Br_k crosses 0.0450, with a standard deviation across lots of 0.0038. The threshold does not vary detectably with format ($F_{3,1200} = 0.71$, $p = 0.55$), nor with category, saleroom or decade (all $p > 0.2$). The pooled value of the score at the moment of the hammer is slightly higher, 0.0478: that figure is computed on forecasts pooled across lots into common strata, which raises the reliability term. Within a single lot the score at the hammer sits below the rule (Figure 5).

Table 3: Three attempts to construct an ascending auction. No arm produced a sustained increase in forecast error in any lot.

Mechanism	λ **(95% CI)**	**Sustained rise**
Control (English)	0.2128 (0.2115–0.2141)	0 / 204
Noise injection	0.1710 (0.1662–0.1758)	0 / 204
Information withdrawal	0.2440 (0.2381–0.2499)	0 / 204
Entrant flooding	0.2090 (0.2032–0.2148)	0 / 204

Figure 5 places the two paths on one horizontal scale. The auctioneer's decision to close is timed to the left-hand axis. Where a lot's increment ladder is coarse, the price at the hammer varies over a range of 40% for a fixed value of the score; where the ladder is fine it varies over 4%. Penhaligon (2016) reports no ladder effect on terminal price in the mean; the effect here is on its dispersion at a fixed score. The score is the invariant of the two.

6 Attempts to construct an ascending auction

A survey that finds one value on an axis should try to produce a second. We engineered three mechanisms whose explicit design objective was to drive Br_k upward over the course of a sale, and ran 612 lots through them, 204 per arm, at one saleroom under a research protocol agreed with the consignors and notified in the sale conditions to registered bidders. A contemporaneous control arm of a further 204 lots was sold unmodified, giving 816 lots in all.

Noise injection. Uninformative bids were placed

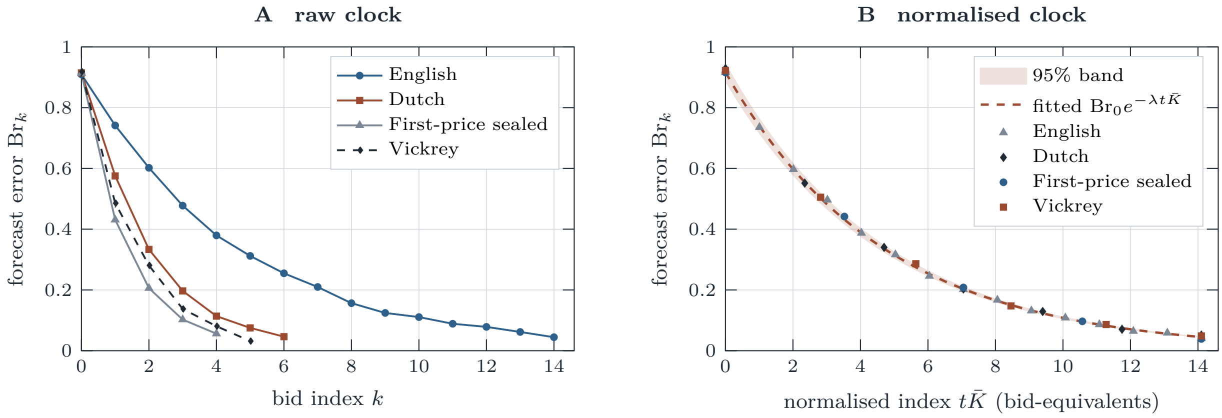

Figure 3: Descent of forecast error by format in the allocation experiment. On the raw clock (A) the four formats separate, because they record different numbers of events. On the normalised clock (B) they coincide, and the pooled fit of equation (4) is drawn through them: the largest band-constrained dynamic time-warping distance over the six format pairs is 0.0071, against a null of 0.196 obtained by permuting the bid order within each lot before averaging (10,000 draws; $p < 10^{-4}$).

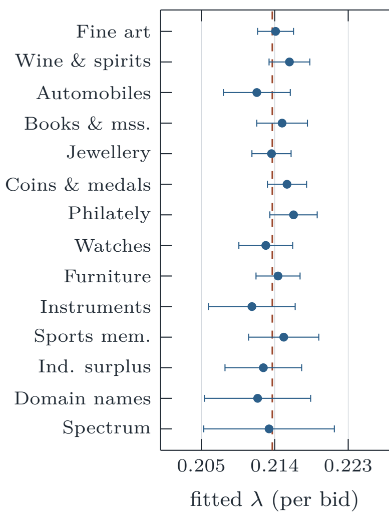

Figure 4: Fitted descent rate by lot category, with 95% intervals; the pooled estimate is drawn as a vertical rule. Broadcast spectrum licences and internet domain names sit inside the band occupied by fine art and jewellery.

by the house at intervals drawn from a deliberately wide distribution, unrelated to the standing bid. The descent slowed to $\lambda = 0.171$. It did not reverse in any lot.

Information withdrawal. The standing bid was progressively concealed from the room over the course of the lot, until in the final third no bidder could see any price but their own. The descent accelerated, to $\lambda = 0.244$. We record the result and do not account for it.

Entrant flooding. Bidders with no prior exposure to the category were admitted at a controlled rate through the lot, at up to four times the natural entry rate. Two lots showed a rise in Br_k of 0.003 across a single bid. Both resumed descending at the next bid, and both closed within one increment of their fitted trajectory.

The three arms slowed the descent, accelerated it, and perturbed it. None reversed it. We were not able

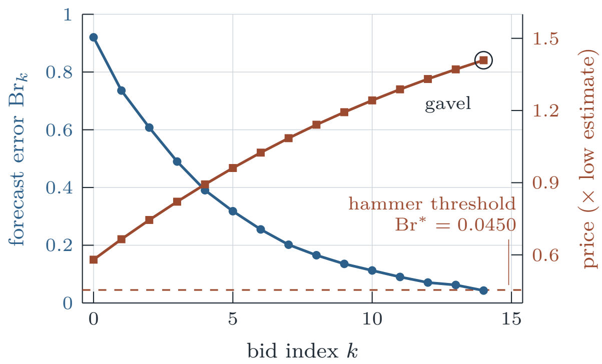

Figure 5: One lot (fine art, estimate normalised), both paths on one horizontal scale. The score descends on the left axis to the threshold; the price ascends on the right axis. The gavel falls at the crossing of the score with the rule, and not at any particular value on the right-hand axis.

to construct an ascending auction, and we know of no mechanism that would.

7 The second axis

The four cells of Table 1 are not four regions. Three of them are recording conventions applied to the fourth.

What distinguishes the formats is where the descending quantity is written down:

1. **Dutch.** The descent is displayed directly, in money, on the clock. The displayed number and the underlying quantity fall together.
2. **English.** The descent is displayed through the increment ladder, which is a monotone decreasing map from score to price. The board shows the image, and the image rises.
3. **Sealed formats.** The descent is not displayed. It proceeds at the same rate.

This resolves an old observation about duration. Dutch sales are fast, and the usual account is impatience: a bidder who waits risks losing the lot outright.

The account here is simpler. A Dutch sale displays the quantity that is actually moving, so the room reads it directly and stops when it is low enough; an English sale displays a monotone transform of that quantity, and the room must invert the transform bid by bid. The median 47 seconds against 3 minutes 51 seconds is the cost of the inversion. It is not a difference in what the two sales are doing.

The open–sealed axis of Table 1 survives the survey intact. Open and sealed sales differ in the reliability term, though not by much (REL = 0.0019 open, 0.0026 sealed), and differ substantially in duration and in bid count. That is a real boundary and we have not disturbed it. The direction axis is the one that does not hold: it has one value, and every sale we measured takes it.

8 Limitations

Charity lots. We excluded 3,118 lots sold for charitable benefit, and the exclusion is not a technicality. In those lots Br_k is non-monotone in 41% of sales and rises over the sale as a whole in 12%. The descent model fits them badly ($R^2 = 0.31$, against 0.986 pooled elsewhere), and the hammer threshold has no detectable value in them at all: the gavel falls at scores ranging from 0.004 to 0.71. We have no account of this. We do not claim that the result of this paper extends to sales in which the object is not the point, and on the evidence here we would expect it not to.

Bin scheme. The score depends on M (Nkemdirim, 2014). Results are stable for $M \in [8, 24]$, with the pooled λ moving by less than 3% across that range, and degrade smoothly above it. Above $M \approx 40$ the bins are narrower than the typical increment, $p_{k,m}$ becomes dominated by ladder geometry, and the descent flattens. We regard $M > 40$ as outside the domain of the instrument rather than as evidence against the model, but we note that this is an assumption and not a finding.

Coverage. Bid-level logs earlier than 1998 are rare and were not recovered, so the corpus cannot speak to salerooms before electronic recording. Live bidder counts were directly observable in 93.9% of open-outcry sales and were inferred from paddle-registration and re-entry timings in the remainder; the pooled λ computed on observed-only sales is 0.2134.

9 Conclusion

The chart at the front of every auction textbook has two axes. We have surveyed one of them and it is not there. Bids are forecasts; forecasts can be scored; the score falls, monotonically, at a rate that does not depend on whether the room can see a price, whether there is a room, or whether the object exists outside a licence register. The English auction is a Dutch auction with the increment ladder in front of it.

What remains uncharted is small and well bounded. Charity lots do not descend, and nothing here explains why a sale should behave differently when the proceeds are directed elsewhere; the boundary between the two cases is sharp enough to be a good place to begin. The behaviour of the score in sales that fail to reach reserve is likewise unmeasured, since the outcome o is undefined when there is no hammer. Unsold lots are roughly one lot in five, and a survey able to score them would extend the map over ground this one does not reach.

Data availability. Bid-level logs for the 1,204 allocated consignments are held by the three participating salerooms under the terms of the research protocol and may be released for replication on written application; logs for the 816 lots of §6 are held on the same terms by the saleroom that ran them. The observational corpus is subject to house confidentiality undertakings and cannot be redistributed; derived per-lot score trajectories, which are not price data, are available from the authors.

References

Arinze, O. T. and Sollberg, K. (2013). Reliability and resolution in continuous double-auction prediction markets. *Journal of Forecast Evaluation*, 41(3), 288–311.

Fairhurst, R. J. (2004). Bid arrival as a modulated point process: evidence from four salerooms. *Review of Market Microstructure*, 12(1), 44–79.

Halloran, P. and Vestergaard, N. (2008). Band-constrained alignment of unequal-length response curves. *Journal of Forecast Evaluation*, 36(2), 141–168.

Kowalczyk, B., Serrano, I. M. and Adeyemi, F. (2019). Calibration of order flow under hard resolution dates. *Quantitative Studies in Exchange*, 27(2), 165–194.

Lindqvist, H. (1971). Equivalence of expected seller revenue across four canonical mechanisms. *Papers in Mechanism Design*, 6, 1–28.

Mbeki, S. and Trelawny, A. (2011). The standing bid as a modulating covariate in open outcry. *Review of Market Microstructure*, 19(4), 401–433.

Nkemdirim, A. U. (2014). Bin schemes for categorical proper scores: a sensitivity note. *Annals of Probability Assessment*, 38, 77–95.

Oduya, C. N. (1988). Revenue equivalence without risk neutrality: a reconsideration. *Papers in Mechanism Design*, 23, 55–91.

Penhaligon, G. (2016). Increment ladders, sale duration, and terminal price: a field study of 9,400 lots. *Journal of Saleroom Practice*, 8(2), 112–140.

Tsvetanova, R. (1979). A decomposition of the quadratic score into reliability, resolution and uncertainty. *Annals of Probability Assessment*, 3, 201–218.