Errata: "Calibeating": Beating Forecasters at Their Own Game*

Dean P. Foster† Sergiu Hart‡

March 19, 2026

While the result of Theorem 6 (in Section 6) in our paper, Foster and Hart (2023)—namely, the existence of a deterministic procedure that is calibeating and continuously calibrated—is correct, its more detailed version, Theorem 12 (in Appendix A.6[1]), is not: the proof that we gave contains an error. We provide here a more refined statement of Theorem 12, together with the corresponding, more extensive, proof.

The error was caused by using the same notation for two distinct concepts of refinement: one is the average within-bin variance of the *actions* a_t (Section 2.1), the other is the average within-bin variance of the *differences* between actions and forecasts $a_t - c_t$ (Appendix A.6). When each bin contains forecasts with the same value, i.e., c_t is constant within the bin, the variance of $a_t - c_t$ equals the variance of a_t, and the two definitions coincide. However, this is not the case for fractional binnings, where a bin usually contains different forecasts. See the end of this errata for details regarding the specific error.

We provide here the necessary corrections for the paper. They are all within Appendix A.6, and consist of a sharpened notation and a revised statement of Theorem 12, together with its expanded proof; the remainder of Appendix A.6—namely, Proposition 13 and its proof—needs no correction beyond using the new notation.

Notation. When we deal with the variance of the differences $a_t - c_t$ rather than the variance of the actions a_t we will add a superscript #. Thus, for the fractional binning

---

*Errata to the published journal version (*Theoretical Economics*, 2023) and to `arXiv` version 2 (2022).

†Department of Statistics, Wharton, University of Pennsylvania, Philadelphia, and Amazon, New York. *e-mail*: `dean@foster.net` *web page*: `http://deanfoster.net`

‡Institute of Mathematics, Department of Economics, and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem. *e-mail*: `hart@huji.ac.il` *web page*: `http://www.ma.huji.ac.il/hart`

[1]In the full paper in `arXiv`, the numbering of the relevant appendices is shifted up by one; Appendix A.6 is thus Appendix A.7 in the full paper.

$(\mathbf{b}, \Pi)$, the variance denoted by $v_t(b,i)$ in the paper is now $v_t^{\#}(b,i)$, i.e.,

$$\begin{aligned} v_t^{\#}(b,i) &:= \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{n_t(b,i)}\right) \|a_s - c_s - e_t(b,i)\|^2 \\ &= \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{n_t(b,i)}\right) \|a_s - c_s - (\bar{a}_t(b,i) - \bar{c}_t(b,i))\|^2 , \end{aligned}$$

where

$$\begin{aligned} \bar{a}_t(b,i) &:= \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{n_t(b,i)}\right) a_s \quad \text{and} \\ \bar{c}_t(b,i) &:= \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{n_t(b,i)}\right) c_s \end{aligned}$$

are the average action and average forecast in the (b,i)-bin, and $e_t(b,i) = \bar{a}_t(b,i) - \bar{c}_t(b,i)$ is the average difference. The refinement score denoted by $\mathcal{R}_t^{\mathbf{b},\Pi}$ in the paper is now $\mathcal{R}_t^{\#;\mathbf{b},\Pi}$, i.e.,

$$\mathcal{R}_t^{\#;\mathbf{b},\Pi} := \sum_{(b,i)\in B\times I} \left(\frac{n_t(b,i)}{t}\right) v_t^{\#}(b,i) = \sum_{(b,i)\in B\times I} \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{t}\right) v_t^{\#}(b,i),$$

and the online refinement score $\widetilde{\mathcal{R}}_t^{\mathbf{b},\Pi}$ is now $\widetilde{\mathcal{R}}_t^{\#;\mathbf{b},\Pi}$, i.e.,

$$\widetilde{\mathcal{R}}_t^{\#;\mathbf{b},\Pi} := \sum_{(b,i)\in B\times I} \sum_{s=1}^{t} \left(\frac{\lambda_s(b,i)}{t}\right) \|a_s - c_s - e_{s-1}(b,i)\|^2 .$$

The same notation is of course used when there is no $\mathbf{b}$, i.e., for Π alone.

The result of Proposition 13 is now written as

$$0 \le \widetilde{\mathcal{R}}_t^{\#;\mathbf{b},\Pi} - \mathcal{R}_t^{\#;\mathbf{b},\Pi} \le o(1).$$

The result of Appendix A.4, namely, the refinement score being monotonically decreasing with respect to refining the binning, is stated for an arbitrary sequence z_t, and so holds for the two kinds of refinement scores; thus, for instance, $\mathcal{R}_t^{\#;\mathbf{b},\Pi} \le \mathcal{R}_t^{\#;\Pi}$ and $\mathcal{R}_t^{\mathbf{b},\Pi} \le \mathcal{R}_t^{\mathbf{b}}$. The general Brier score decomposition of Appendix A.5 for fractional binnings is now written as $\mathcal{B}_t^{\mathbf{c}} = \mathcal{R}_t^{\#;\Pi} + \mathcal{K}_t^{\Pi}$.

We now state the corrected version of Theorem 12.

Theorem 12 (corrected) *Let B be a finite set. Then there exists a* deterministic $\mathbf{b}$-*based forecasting procedure ζ that is B-calibeating and is continuously calibrated. Specifically: first, for every continuous binning Π there is a deterministic* $\mathbf{b}$*-based forecasting* *procedure* $\zeta \equiv \zeta_\Pi$ *such that*

$$\mathcal{K}_t^{\mathbf{b},\Pi} = \mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\mathbf{b},\Pi} \leq o(1); \tag{34}$$

and second, there is a continuous binning Π^* *for which (34) implies that the corresponding procedure* ζ_{Π^*} *is B-calibeating, i.e.,*

$$\mathcal{B}_t^{\mathbf{c}} \leq \mathcal{R}_t^{\mathbf{b}} + o(1),$$

and is continuously calibrated. All the above inequalities hold as $t \to \infty$ *uniformly over all sequences* $\mathbf{a}$ *and* $\mathbf{b}$.

The statement of Theorem 12 in Foster and Hart (2023) claimed incorrectly that the result holds for Π^* equal to the continuous binning Π_0 given by Proposition 3 in Foster and Hart (2021); as we will show below, we need Π^* to be a certain "expansion" of Π_0.

The proof relies on a result that shows that the two kinds of refinements, $\mathcal{R}^\#$ and $\mathcal{R}$, are close when the binning is "local," in the sense that the forecasts within each bin are almost constant.

A weight function $w : C \to [0, 1]$ is δ-*local* (for $\delta > 0$) if all points with positive weight (the *support* of w) lie in an open ball of radius δ; i.e., $\{c \in C : w(c) > 0\} \subseteq B(y; \delta)$ for some $y \in C$. A fractional binning $\Pi = (w_i)_{i \in I}$ is δ-*local* if all the weight functions w_i are δ-local. Such a fractional binning $\Pi \equiv \Pi_\delta$ can be obtained, for instance, by using the so-called "δ-tent functions" based on a δ-grid: let $C_\delta = \{y^i\}_{i \in I}$ be a finite δ-grid of C, and for each $i \in I$ let $w_i(c) := \Lambda(c, y^i)/\sum_{j \in I} \Lambda(c, y_j)$, where $\Lambda(c, y) := [\delta - \|c - y\|]_+$; moreover, this construction yields a continuous binning (because the functions $\Lambda(\cdot, y^i)$ are continuous, and their sum is positive, and thus bounded away from 0 on the compact set C).

Lemma *Let* Π *be a* δ*-local fractional binning. Then*

$$\left| \mathcal{R}_t^{\#;\Pi} - \mathcal{R}_t^{\Pi} \right| < 2\delta + \delta^2.$$

Proof. Let $\Pi = (w_i)_{i \in I}$, and for each $i \in I$ let $y^i \in C$ be such that $B(y^i; \delta)$ contains the support of w_i. Thus, all forecasts c_s counted in bin i (i.e., with $\lambda_s(i) = w_i(c_s) > 0$) satisfy $\|c_s - y^i\| < \delta$.

Given random variables X and Y such that $\|X\| \leq 1$ and $\|Y\| < \delta$, and hence $\mathbb{V}ar[X] \leq 1$ and $\mathbb{V}ar[Y] < \delta^2$, we have

$$|\mathbb{V}ar[X - Y] - \mathbb{V}ar[X]| \leq 2\sqrt{\mathbb{V}ar[X]\,\mathbb{V}ar[Y]} + \mathbb{V}ar[Y] < 2\delta + \delta^2.$$

Applying this to $X = a_s$ and $Y = c_s - y^i$ (since y^i is a constant it does not affect the variance) yields

$$\left| v_t^\#(i) - v_t(i) \right| < 2\delta + \delta^2,$$

where

$$\begin{aligned} v_t^{\#}(i) &= \sum_{s=1}^{t} \left(\frac{\lambda_s(i)}{n_t(i)} \right) \|a_s - c_s - (\bar{a}_t(i) - \bar{c}_t(i))\|^2 \quad \text{and} \\ v_t(i) &= \sum_{s=1}^{t} \left(\frac{\lambda_s(i)}{n_t(i)} \right) \|a_s - \bar{a}_t(i)\|^2 \end{aligned}$$

are the variances in bin i of the differences $a_s - c_s$, and of the actions a_s, respectively. Averaging over $i \in I$, with weights $n_t(i)/t$, gives the result. □

We now prove the corrected Theorem 12. Part (i) is the original proof (for an arbitrary Π), and part (ii) is new.

Proof of Theorem 12 (corrected). (i) Let $\Pi = (w_i)_{i\in I}$ be a continuous binning. At time t, given $\mathbf{a}_{t-1}, \mathbf{c}_{t-1}$, and $\mathbf{b}_t$, applying the outgoing fixed point result, specifically, Theorem 10 (D), to the continuous function $c \longmapsto c + \sum_{i\in I} w_i(c) e_{t-1}(b_t, i)$, yields $c_t \in C$ such that

$$\begin{aligned} \|a_t - c_t\|^2 &\leq \left\| a_t - c_t - \sum_{i\in I} w_i(c_t) e_{t-1}(b_t, i) \right\|^2 \\ &\leq \sum_{i\in I} w_i(c_t) \|a_t - c_t - e_{t-1}(b_t, i)\|^2 \end{aligned}$$

for every $a_t \in A$ (the second inequality is by the convexity of $\|\cdot\|^2$). Averaging over t gives $\mathcal{B}_t^{\mathbf{c}} \leq \widetilde{\mathcal{R}}_t^{\#;\mathbf{b},\Pi}$, and thus (34) by Proposition 13 together with the decomposition of Appendix A.5 for the fractional binning $(\mathbf{b}, \Pi)$.

(ii) We construct Π^* as follows. Let $\Pi_0 = (w_i^0)_{i\in I_0}$ be the continuous binning given by Proposition 3 in Foster and Hart (2021). For each $n \geq 1$ let $\Pi_n = (w_i^n)_{i\in I_n}$ be a δ_n-local continuous binning, where $0 < \delta_n < 1$ and $\delta_n \to 0$ as $n \to \infty$; assume that the indexing sets I_n for all $n \geq 0$ are taken to be disjoint. The collection of all these weight functions, with w_i^n rescaled by a factor of $1/2^{n+1}$, yields a continuous binning, which we denote by Π^*; i.e., $\Pi^* = (w_i^*)_{i\in I^*}$ with $I^* := \cup_{n\geq 0} I_n$ and $w_i^* := (1/2^{n+1}) w_i^n$ for each $i \in I_n$ and $n \geq 0$ (indeed, every w_i^* is continuous, and $\sum_{i\in I} w_i^* = \sum_{n\geq 0} \sum_{i\in I_n} (1/2^{n+1}) w_i^n = \sum_{n\geq 0} (1/2^{n+1}) \mathbf{1} = \mathbf{1}$, because $\sum_{i\in I_n} w_i^n = \mathbf{1}$ for each n).

Consider now the joint fractional binning $(\mathbf{b}, \Pi^*)$, and the corresponding refinement score $\mathcal{R}_t^{\#;\mathbf{b},\Pi^*}$. Separating the sum over all $i \in I$ into sums over $i \in I_n$ for all $n \geq 0$, and noting that rescaling a weight function does not affect the within-bin *relative* weights $\lambda_s(b,i)/n_t(b,i) = \mathbf{1}_b(b_s) \cdot w_i(c_s)/\sum_{r\leq t} w_i(c_r)$ that are used to compute the bin averages and variances, we get

$$\mathcal{R}_t^{\#;\mathbf{b},\Pi^*} = \sum_{n\geq 0} \frac{1}{2^{n+1}} \mathcal{R}_t^{\#;\mathbf{b},\Pi_n}.$$

The procedure ζ_{Π^*} constructed in (i) for Π^* yields, by (34),

$$\mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\mathbf{b},\Pi^*} = \sum_{n\geq 0} \frac{1}{2^{n+1}} \left(\mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\mathbf{b},\Pi_n} \right) \leq o(1).$$

By the decomposition of Appendix A.5 for $(\mathbf{b}, \Pi_n)$, each term in the above sum equals $\mathcal{K}_t^{\mathbf{b},\Pi_n}$, and is thus nonnegative, and so

$$\mathcal{K}_t^{\mathbf{b},\Pi_n} = \mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\mathbf{b},\Pi_n} \leq o(1) \quad \text{for every } n \geq 0. \tag{E1}$$

For $n = 0$, (E1) implies that

$$\mathcal{K}_t^{\Pi_0} = \mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\Pi_0} \leq \mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\#;\mathbf{b},\Pi_0} \leq o(1),$$

where we have used the decomposition of Appendix A.5 for Π_0, and $\mathcal{R}_t^{\#;\mathbf{b},\Pi_0} \leq \mathcal{R}_t^{\#;\Pi_0}$ by the refining monotonicity of the refinement score (Appendix A.4). Thus, by Proposition 3 in Foster and Hart (2021), the procedure is continuously calibrated.

Next, for each $n \geq 1$, (E1) implies that

$$\mathcal{B}_t^{\mathbf{c}} \leq \mathcal{R}_t^{\#;\mathbf{b},\Pi_n} + o(1) \leq \mathcal{R}_t^{\mathbf{b},\Pi_n} + 3\delta_n + o(1) \leq \mathcal{R}_t^{\mathbf{b}} + 3\delta_n + o(1),$$

where we have used $\mathcal{R}_t^{\#;\mathbf{b},\Pi_n} \leq \mathcal{R}_t^{\mathbf{b},\Pi_n} + 2\delta_n + \delta_n^2 < \mathcal{R}_t^{\mathbf{b},\Pi_n} + 3\delta_n$ by the above lemma applied to the fractional binning $(\mathbf{b}, \Pi_n)$, which is δ_n-local, and $\mathcal{R}_t^{\mathbf{b},\Pi_n} \leq \mathcal{R}_t^{\mathbf{b}}$ (again by Appendix A.4). Therefore $\mathcal{B}_t^{\mathbf{c}} \leq \mathcal{R}_t^{\mathbf{b}} + 4\delta_n$ for all t large enough; since $\delta_n \to 0$, this proves that $\mathcal{B}_t^{\mathbf{c}} - \mathcal{R}_t^{\mathbf{b}} \leq o(1)$. □

We can now pinpoint the error: it lies in the inequality $\mathcal{R}_t^{\mathbf{b},\Pi_0} \leq \mathcal{R}_t^{\mathbf{b}}$ in the paper, which we took to mean (in the current notation) $\mathcal{R}_t^{\#;\mathbf{b},\Pi_0} \leq \mathcal{R}_t^{\mathbf{b}}$, whereas what we showed was $\mathcal{R}_t^{\#;\mathbf{b},\Pi_0} \leq \mathcal{R}_t^{\#;\mathbf{b}}$. We have resolved this by supplementing Π_0 with continuous binnings Π_δ for which all forecasts within the same bin are close to one another, and so the variance of the differences $a_t - c_t$ is close to the variance of the actions a_t, and thus $\mathcal{R}_t^{\mathbf{b},\Pi_\delta}$ is close to $\mathcal{R}_t^{\#;\mathbf{b},\Pi_\delta}$.

Misprints

- Page 1441, 5th line from the bottom: replace "preset paper" with "present paper."
- Page 1450, 7th line after formula (6): replace "$\phi : B \to B$" with "$\phi : B \to \Delta(A)$."
- Page 1457, Theorem 7 (iii): replace "(δ, B_n)-calibeating" with "B_n-calibeating."
- Page 1472, displayed formula in Remark (b): replace $\bar{a}_{t-1}^n$ with $\bar{a}_{i,t-1}^n$ (add the missing subscript i).

References

Foster, D. P. and S. Hart (2021), "Forecast Hedging and Calibration," *Journal of Political Economy* 129, 3447–3490.

Foster, D. P. and S. Hart (2023), "'Calibeating': Beating Forecasters at Their Own Game," *Theoretical Economics* 18, 1441–1474.

- Full version (2022): `http://arxiv.org/abs/2209.04892v2`

- Addendum (2024): `http://ma.huji.ac.il/hart/papers/calib-beat-add.pdf`