

813 A Proof of the theoretical results

814 In this section, we restate each theoretical result and provide proofs.

815 A.1 Proof of Proposition 1

816 **Proposition 1.** *The combined estimator $\hat{F}_{1:t} = \sum_{\tau=1}^t \bar{\alpha}_\tau \hat{F}_\tau$ is unbiased: $\mathbb{E}[\hat{F}_{1:t}] = F(\Omega)$.*

817 *Proof.* We first show that each individual estimator \hat{F}_τ is unbiased.

$$\begin{aligned} \mathbb{E}_{\mathcal{D}_\tau, s_\tau}[\hat{F}_\tau] &= \mathbb{E}_{\mathcal{D}_\tau, s_\tau} \left[F(D_\tau) + \frac{f(s_\tau)}{q_\tau(s_\tau)} \right] \\ &= \mathbb{E}_{\mathcal{D}_\tau} \left[F(D_\tau) + \mathbb{E}_{s_\tau} \left[\frac{f(s_\tau)}{q_\tau(s_\tau)} \right] \right] \\ &= \mathbb{E}_{\mathcal{D}_\tau} \left[F(D_\tau) + \sum_{s_\tau \in \Omega \setminus \mathcal{D}_\tau} q_\tau(s_\tau) \frac{f(s_\tau)}{q_\tau(s_\tau)} \right] \\ &= \mathbb{E}_{\mathcal{D}_\tau} [F(D_\tau) + F(\Omega \setminus \mathcal{D}_\tau)] \\ &= F(\Omega). \end{aligned}$$

818 Therefore, for the combined estimator

$$\begin{aligned} \mathbb{E}[\hat{F}_{1:t}] &= \mathbb{E} \left[\sum_{\tau=1}^t \bar{\alpha}_\tau \hat{F}_\tau \right] \\ &= \sum_{\tau=1}^t \bar{\alpha}_\tau \mathbb{E}[\hat{F}_\tau] \\ &= \sum_{\tau=1}^t \bar{\alpha}_\tau F(\Omega) \\ &= F(\Omega). \end{aligned}$$

819

□

820 A.2 Proof of proposition 2

821 **Proposition 2.** *For any $1 \leq \tau < r \leq t$, $\text{Cov}(\hat{F}_\tau, \hat{F}_r) = 0$.*

Proof.

$$\begin{aligned} \text{Cov}[\hat{F}_\tau, \hat{F}_r] &= \mathbb{E}[(\hat{F}_\tau - F(\Omega))(\hat{F}_r - F(\Omega))] \\ &= \mathbb{E}_{\mathcal{D}_r}[(\hat{F}_\tau - F(\Omega)) \mathbb{E}_{s_r}[\hat{F}_r - F(\Omega)]] \\ &= \mathbb{E}_{\mathcal{D}_r} \left[(\hat{F}_\tau - F(\Omega)) \sum_{s_r \in \Omega \setminus \mathcal{D}_r} q_r(s_r) \left(F(\mathcal{D}_r) + \frac{f(s_r)}{q_r(s_r)} - F(\Omega) \right) \right] \\ &= \mathbb{E}_{\mathcal{D}_r} \left[(\hat{F}_\tau - F(\Omega)) \left(F(\mathcal{D}_r) + \sum_{s_r \in \Omega \setminus \mathcal{D}_r} q_r(s_r) \frac{f(s_r)}{q_r(s_r)} - F(\Omega) \right) \right] \\ &= \mathbb{E}_{\mathcal{D}_r} [(\hat{F}_\tau - F(\Omega)) \cdot 0] \\ &= 0. \end{aligned}$$

822

□

823 A.3 Proof of proposition 3

824 **Proposition 3.** *If there exists constants $0 < A \leq B$ such that for any $s \in \Omega$, $A \leq f(s) \leq B$ and*
 825 *$A \leq g(s) \leq B$, then there exists a constant $C > 0$ such that $\text{Var}[\hat{F}_\tau] \leq C(N - \tau)(N - \tau + 1)$.*

826 *Proof.* To start, we first bound the sampling distribution implied by the detector measurements $g(s)$.
 827 For unit $s \in \Omega \setminus \mathcal{D}_\tau$, the sampling distribution is $q_\tau(s) = \frac{g(s)}{\sum_{s' \in \Omega \setminus \mathcal{D}_\tau} g(s')}$. By the lower and upper
 828 bounds, we can see that $\frac{A}{(N - \tau + 1)B} \leq q_\tau(s) \leq \frac{B}{(N - \tau + 1)A}$. Also, we can bound $F(\Omega) - F(\mathcal{D}_\tau) =$
 829 $\sum_{s \in \Omega \setminus \mathcal{D}_\tau} f(s)$ with the same technique, which is $(N - \tau + 1)A \leq F(\Omega) - F(\mathcal{D}_\tau) \leq (N - \tau + 1)B$.
 830 Next, we write the variance of \hat{F}_τ in terms of expectations.

$$\begin{aligned} \text{Var}[\hat{F}_\tau] &= \mathbb{E}_{\mathcal{D}_\tau, s_\tau} \left[\left(F(\mathcal{D}_\tau) + \frac{f(s_\tau)}{q_\tau(s_\tau)} - F(\Omega) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{D}_\tau, s_\tau} \left[\frac{f(s_\tau)^2}{q_\tau(s_\tau)^2} + (F(\mathcal{D}_\tau) - F(\Omega))^2 + 2(F(\mathcal{D}_\tau) - F(\Omega)) \frac{f(s_\tau)}{q_\tau(s_\tau)} \right] \\ &\leq \mathbb{E}_{\mathcal{D}_\tau, s_\tau} \left[\frac{(N - \tau + 1)^2 B^4}{A^2} + (N - \tau + 1)^2 B^2 - 2(N - \tau + 1)^2 \frac{A^3}{B} \right] \\ &= (N - \tau + 1)^2 \frac{B^3(B^2 + A^2) - 2A^5}{A^2 B}. \end{aligned}$$

831 When $\tau < N$, we further have

$$\begin{aligned} \text{Var}[\hat{F}_\tau] &= (N - \tau + 1)^2 \frac{B^3(B^2 + A^2) - 2A^5}{A^2 B} \\ &= (N - \tau)(N - \tau + 1) \frac{N - \tau + 1}{N - \tau} \frac{B^3(B^2 + A^2) - 2A^5}{A^2 B} \\ &\leq (N - \tau)(N - \tau + 1) \frac{2B^3(B^2 + A^2) - 4A^5}{A^2 B}. \end{aligned}$$

832 Also note that

$$\begin{aligned} \text{Var}[\hat{F}_N] &= \mathbb{E}_{\mathcal{D}_N, s_N} \left[\left(F(\mathcal{D}_N) + \frac{f(s_N)}{q_N(s_N)} - F(\Omega) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{D}_N, s_N} \left[\left(F(\mathcal{D}_N) + \frac{f(s_N)}{1} - F(\Omega) \right)^2 \right] \\ &= \mathbb{E}_{\mathcal{D}_N, s_N} \left[(F(\Omega) - F(\Omega))^2 \right] = 0. \end{aligned}$$

833 This means that for all $\tau \leq N$, $\text{Var}[\hat{F}_\tau] \leq C(N - \tau)(N - \tau + 1)$, where $C = \frac{2B^3(B^2 + A^2) - 4A^5}{A^2 B}$. \square

834 A.4 Proof of proposition 4

835 **Proposition 4.** *If $\text{Var}[\hat{F}_\tau] \propto \tau^{-y}/w_\tau$, where $y \in [0, 1]$, and w_τ are non-decreasing, then the*
 836 *estimate $\hat{F}_{1:t}^{\text{comb}} = \sum_{\tau=1}^t \alpha_\tau^{\text{comb}} \hat{F}_\tau$, where $\alpha_\tau^{\text{comb}} \propto w_\tau \sqrt{\tau}$ and $\sum_{\tau=1}^t \alpha_\tau^{\text{comb}} = 1$, satisfy*

$$\sup_{t \geq 1} \sup_{y \in [0, 1]} \frac{\text{Var}[\hat{F}_{1:t}^{\text{comb}}]}{\text{Var}_{\text{opt}}(y, t)} \leq \frac{9}{8},$$

837 where $\text{Var}_{\text{opt}}(y, t)$ is the estimation variance when the estimators are weighted proportional to
 838 inverse of ground-truth variances.

839 *Proof.* Denote the variance of combined estimator weighted by $\alpha_\tau^{(x)} \propto w_\tau \tau^x$ by $\text{Var}_x(y, t)$, where
 840 $x \in [0, 1]$. An observation is that $\text{Var}[\hat{F}_{1:t}^{\text{comb}}] = \text{Var}_{0.5}(y, t)$, and $\text{Var}_{\text{opt}}(y, t) = \text{Var}_y(y, t)$. So we

841 shall bound $\frac{\text{Var}_{0.5}(y,t)}{\text{Var}_y(y,t)}$. Note that for any $x \in [0, 1]$,

$$\begin{aligned}\text{Var}_x(y, t) &\propto \frac{\sum_{\tau=1}^t w_{\tau}^2 \tau^{2x} \cdot \tau^{-y} / w_{\tau}}{\left(\sum_{\tau=1}^t w_{\tau} \tau^x\right)^2} \\ &= \frac{\sum_{\tau=1}^t w_{\tau} \tau^{2x-y}}{\left(\sum_{\tau=1}^t w_{\tau} \tau^x\right)^2}.\end{aligned}$$

842 Therefore, let

$$l_x(y, t) = \frac{\text{Var}_x(y, t)}{\text{Var}_y(y, t)} = \frac{\left(\sum_{\tau=1}^t w_{\tau} \tau^{2x-y}\right) \left(\sum_{\tau=1}^t w_{\tau} \tau^y\right)}{\left(\sum_{\tau=1}^t w_{\tau} \tau^x\right)^2},$$

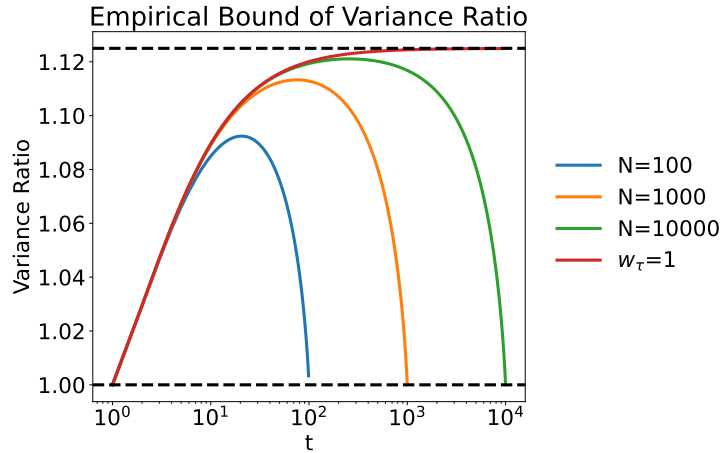
843 we have

$$\frac{\partial^2 l_x(y, t)}{\partial y^2} = \frac{\sum_{\tau=1}^t \sum_{\tau'=1}^t w_{\tau} w_{\tau'} \tau^{2x-y} \tau'^y (\log \tau - \log \tau')^2}{\left(\sum_{\tau=1}^t w_{\tau} \tau^x\right)^2} > 0.$$

844 So $l_x(y, t)$ is strictly convex over y . Also note that $l_x(y, t)$ is symmetric around $y = x$. So
 845 $\sup_{y \in [0,1]} l_x(y, t) = \begin{cases} l_x(1, t) & x \leq 1/2 \\ l_x(0, t) & x \geq 1/2 \end{cases}$. In the combined estimator we chose $x = 0.5$, which leads
 846 to the supremum $l_{0.5}(1, t)$ for any t . Furthermore, $x = 0.5$ is the best assumption for the mixing
 847 weights. The derivation is the same as Owen and Zhou [24] and we omit it here. In such case

$$\sup_{y \in [0,1]} l_{0.5}(y, t) = l_{0.5}(1, t) = \frac{\left(\sum_{\tau=1}^t w_{\tau}\right) \left(\sum_{\tau=1}^t w_{\tau} \tau\right)}{\left(\sum_{\tau=1}^t w_{\tau} \tau^{0.5}\right)^2} = \frac{\sum_{\tau=1}^t \frac{w_{\tau}}{\left(\sum_{\tau'=1}^t w_{\tau'}\right)^{\tau}}}{\left(\sum_{\tau=1}^t \frac{w_{\tau}}{\left(\sum_{\tau'=1}^t w_{\tau'}\right)^{0.5}}\right)^2}.$$

848 We can plot this specific function for a few different choices of w_{τ} and t ,



849 The lines labeled $N = x$ correspond to the combined weighting scheme proposed in our main method
 850 (see Section 4), and $w_{\tau} = 1$ is a uniform weighting. We can see empirically that these are all bounded
 851 above by $\frac{9}{8}$.

852 To prove this bound, we define the following weights,

$$\hat{w}_{\tau} = \frac{w_{\tau}}{\left(\sum_{\tau'=1}^t w_{\tau'}\right)}, \quad \beta_{\tau} = (\hat{w}_{\tau} - \hat{w}_{\tau-1})(t - \tau + 1).$$

853 Since w_τ are non-decreasing, $\hat{w}_\tau - \hat{w}_{\tau-1} \geq 0$ and thus $\beta_\tau \geq 0$. We can also cancel out terms of a
 854 telescoping series to see that $\sum_{\tau=1}^t \beta_\tau = \sum_{\tau=1}^t \hat{w}_\tau = 1$. Therefore both \hat{w}_τ and β_τ are non-negative
 855 weights that sum to 1. Note we define $\hat{w}_0 = 0$.

856 Plugging these in, and borrowing a trick from Gabriel's Staircase, we can rewrite as,

$$\frac{\sum_{\tau=1}^t \frac{w_\tau}{(\sum_{\tau'=1}^t w_{\tau'})^\tau} \tau}{\left(\sum_{\tau=1}^t \frac{w_\tau}{(\sum_{\tau'=1}^t w_{\tau'})^{\tau^{0.5}}} \tau^{0.5} \right)^2} = \frac{\sum_{\tau=1}^t \hat{w}_\tau \tau}{\left(\sum_{\tau=1}^t \hat{w}_\tau \tau^{0.5} \right)^2} = \frac{\sum_{\tau=1}^t \beta_\tau \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'}{\left(\sum_{\tau=1}^t \beta_\tau \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'^{0.5} \right)^2}.$$

857 We can bound this expression by considering the supremum over all possible positive weights λ_τ ,

$$\begin{aligned} \frac{\sum_{\tau=1}^t \beta_\tau \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'}{\left(\sum_{\tau=1}^t \beta_\tau \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'^{0.5} \right)^2} &\leq \sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{\sum_{\tau=1}^t \frac{\lambda_\tau}{\sum_{\tau'=1}^t \lambda_{\tau'}} \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{\sum_{\tau'=1}^t \lambda_{\tau'}} \frac{1}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'^{0.5} \right)^2} \\ &= \sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{\left(\sum_{\tau=1}^t \lambda_\tau \right) \left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} \sum_{\tau'=\tau}^t \tau' \right)}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} \sum_{\tau'=\tau}^t \tau'^{0.5} \right)^2}. \end{aligned}$$

858 We can simplify by using the closed form expression for sum of consecutive integers and a lower
 859 bound for the sum of consecutive square roots from Owen and Zhou [24],

$$\begin{aligned} &\leq \sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{\left(\sum_{\tau=1}^t \lambda_\tau \right) \left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} \frac{(t+\tau)(t-\tau+1)}{2} \right)}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} \frac{(t+0.5)^{1.5} - (\tau-0.5)^{1.5}}{1.5} \right)^2} \\ &= \sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{9}{8} \frac{\left(\sum_{\tau=1}^t \lambda_\tau \right) \left(\sum_{\tau=1}^t \lambda_\tau (t+\tau) \right)}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} ((t+0.5)^{1.5} - (\tau-0.5)^{1.5}) \right)^2}. \end{aligned} \quad (8)$$

860 Define this last expression inside the supremum as $f(\lambda, t)$. Taking the derivative with respect to each
 861 λ_n ,

$$\begin{aligned} \frac{\partial f(\lambda, t)}{\partial \lambda_n} &= \frac{\left(\sum_{\tau=1}^t \lambda_\tau (t+\tau) \right) + (t+n) \left(\sum_{\tau=1}^t \lambda_\tau \right)}{\left(\sum_{\tau=1}^t \lambda_\tau \frac{1}{t-\tau+1} ((t+0.5)^{1.5} - (\tau-0.5)^{1.5}) \right)^2} \\ &\quad - \frac{2 \left(\sum_{\tau=1}^t \lambda_\tau \right) \left(\sum_{\tau=1}^t \lambda_\tau (t+\tau) \right) \left(\frac{1}{t-n+1} ((t+0.5)^{1.5} - (n-0.5)^{1.5}) \right)}{\left(\sum_{\tau=1}^t \lambda_\tau \frac{1}{t-\tau+1} ((t+0.5)^{1.5} - (\tau-0.5)^{1.5}) \right)^3}. \end{aligned}$$

862 Setting this equal to 0 and rearranging,

$$\frac{\partial f(\lambda, t)}{\partial \lambda_n} = 0$$

863

$$\frac{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{\sum_{\tau'=1}^t \lambda_{\tau'}} \tau \right) + (2t+n)}{\left(\frac{1}{t-n+1} ((t+0.5)^{1.5} - (n-0.5)^{1.5}) \right)} = \frac{2 \left(\sum_{\tau=1}^t \lambda_\tau (t+\tau) \right)}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t-\tau+1} ((t+0.5)^{1.5} - (\tau-0.5)^{1.5}) \right)}. \quad (9)$$

864 Note the right hand side does not depend on n , the index of the partial derivative. Thus at a critical
 865 point, the left hand side will equal the same value for all n . We can define this left side as a function
 866 $g(n, t, \lambda)$.

867 To simplify, we make the following substitutions:

$$\begin{aligned} c_1(t, \lambda) &= \left(\sum_{\tau=1}^t \frac{\lambda_\tau}{\sum_{\tau'=1}^t \lambda_{\tau'}} \tau \right) - 0.5 \\ u(n) &= \sqrt{n-0.5} \\ w(t) &= \sqrt{t+0.5} \end{aligned}$$

868 Since $n \geq 1$, $t \geq 1$, and $n \leq t$, we can note that $u(n) > 0$, $w(t) > 0$, $u(n) < w(t)$, and
 869 $0.5 \leq c_1(t, \lambda) \leq t - 0.5$. This alleviates concerns over dividing by zero going forward.

870 Applying these to $g(n, t, \lambda)$,

$$g(n, t, \lambda) = \frac{c_1(t, \lambda) + 2w(t)^2 + u(n)^2}{\frac{w(t)^3 - u(n)^3}{w(t)^2 - u(n)^2}} = \frac{(w(t) + u(n))(c_1(t, \lambda) + 2w(t)^2 + u(n)^2)}{w(t)^2 + w(t)u(n) + u(n)^2}.$$

871 Next we can consider the partial derivative over n and set it equal to zero, noting that $u'(n) = \frac{1}{2u(n)}$,

$$\frac{\partial g(n, t, \lambda)}{\partial n} = \frac{u(n)(u(n)^3 + 2w(t)u(n)^2 + (2w(t)^2 - c_1(t, \lambda))u(n) - 2w(t)(w(t)^2 + c_1(t, \lambda)))}{2u(n)^2 + w(t)u(n) + w(t)^2} = 0.$$

872 This is a cubic equation over $u(n)$. Looking at the coefficients, 1 and $2w(t)$ are both positive. Making
 873 a substitution and using a bound on c_1 , $2w(t)^2 - c_1(t, \lambda) \geq 2t + 1 - t + 0.5 = t + 1.5 > 0$. Clearly
 874 the last term is negative. By Descartes' rule of signs, the coefficients of this polynomial have 1 sign
 875 change and thus at most 1 positive real root over $u(n)$. Since $u(n)$ is monotonic over n and positive
 876 for all $n \geq 1$, we see that there exists at most one n such that $\frac{\partial g(n, t, \lambda)}{\partial n} = 0$, for any given t and λ .
 877 From this we can deduce that the equation $g(n, t, \lambda) = h(t, \lambda)$ for any function $h(t, \lambda)$ that does not
 878 depend on n has at most 2 solutions over n .

879 Applying this to Eq. 9 and noting that the left hand side is equal to $g(n, t, \lambda)$, we can conclude that
 880 this equation can have no more than 2 solutions over n . Thus the partial derivative of $f(\lambda, t)$ with
 881 respect to λ can have at most 2 components with value 0 at the same time. We can also observe
 882 $f(c\lambda, t) = f(\lambda, t)$ for all $c \in \mathbb{R}^+$, thus the supremum must occur at some finite point. Since we're
 883 optimizing over the positive orthant, that finite point must have all non-zero components with gradient
 884 zero. Thus there can be at most 2 non-zero components of λ at the supremum. Continuing from Eq.
 885 8, we now have the upper bound,

$$\sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{9}{8} \frac{\left(\sum_{\tau=1}^t \lambda_\tau\right) \left(\sum_{\tau=1}^t \lambda_\tau(t + \tau)\right)}{\left(\sum_{\tau=1}^t \frac{\lambda_\tau}{t - \tau + 1} ((t + 0.5)^{1.5} - (\tau - 0.5)^{1.5})\right)^2} \\ = \sup_{\lambda \in \mathbb{R}_{\geq 0}^t} \frac{9}{8} \frac{(\lambda_i + \lambda_j)(\lambda_i(t + i) + \lambda_j(t + j))}{\left(\frac{\lambda_i}{t - i + 1} ((t + 0.5)^{1.5} - (i - 0.5)^{1.5}) + \frac{\lambda_j}{t - j + 1} ((t + 0.5)^{1.5} - (j - 0.5)^{1.5})\right)^2}. \quad (10)$$

886 We will denote this function inside the supremum as z . We can replace λ_i and λ_j with variables x, y
 887 and make similar substitutions as before,

$$u(i) = \sqrt{i - 0.5} \\ v(j) = \sqrt{j - 0.5} \\ w(t) = \sqrt{t + 0.5}$$

888 For more compact notation we will ignore arguments going forward. Making these substitutions
 889 yields,

$$z(x, y, i, j, t) = \frac{9}{8} \frac{(x + y)(x(w^2 + u^2) + y(w^2 + v^2))}{\left(\frac{x(w^3 - u^3)}{w^2 - u^2} + \frac{y(w^3 - v^3)}{w^2 - v^2}\right)^2}.$$

890 Using a computer algebra system, we can expand this as,

$$z(x, y, i, j, t) = \frac{9}{8} \left(1 - \frac{p}{q}\right) \\ p = -u^4 v^2 xy - 2u^4 v w xy - u^4 w^2 xy + 2u^3 v^3 xy + 2u^3 v^2 w xy - u^2 v^4 xy + 2u^2 v^3 w xy + \\ u^2 v^2 w^2 x^2 + 4u^2 v^2 w^2 xy + u^2 v^2 w^2 y^2 + 2u^2 v w^3 x^2 + 2u^2 v w^3 xy + u^2 w^4 x^2 + \\ u^2 w^4 xy - 2u v^4 w xy + 2u v^2 w^3 xy + 2u v^2 w^3 y^2 - v^4 w^2 xy + v^2 w^4 xy + v^2 w^4 y^2 \\ q = (x(w^2 + wu + u^2)(w + v) + y(w^2 + wv + v^2)(w + u))^2.$$

891 By definition $x, y \geq 0$ and $1 \leq i < j \leq t$, so $u, w, v > 0$, $u < w$, and $v < w$. Thus we can see that
 892 $q > 0$. Rearranging terms,

$$\begin{aligned} p = & u^3 v^2 (2w - u) xy + 2u^2 vw (w^2 - u^2) xy + u^2 w^2 (w^2 - u^2) xy \\ & + u^2 v^2 (4w^2 - v^2) xy + 2uv^2 w (w^2 - v^2) xy + v^2 w^2 (w^2 - v^2) xy \\ & + 2u^3 v^3 xy + 2u^2 v^3 w xy + u^2 v^2 w^2 x^2 + u^2 v^2 w^2 y^2 + 2u^2 v w^3 x^2 \\ & + u^2 w^4 x^2 + 2uv^2 w^3 y^2 + v^2 w^4 y^2. \end{aligned}$$

893 Since $w > u$ and $w > v$ we can see that p is a sum of positive terms so $p > 0$. Putting this together
 894 we have $\frac{p}{q} > 0$. Therefore continuing from Eq. 10,

$$= \sup_{t \geq 1, x, y \geq 0, 1 \leq i < j \leq t} z(x, y, i, j, t) = \sup_{t \geq 1, x, y \geq 0, 1 \leq i < j \leq t} \frac{9}{8} \left(1 - \frac{p}{q}\right) \leq \frac{9}{8}.$$

895 This proves our bound. In conclusion, we first show that the variance ratio is bounded by another
 896 ratio involving 2 weighted averages. We then use the fact that the weights are non-decreasing to
 897 transform the expressions into a weighted average of unweighted averages. Using closed forms and
 898 approximations we find a simpler upper bound. We consider the supremum of this expression over
 899 all possible weights, and observe that the maximum can only occur with at most 2 non-zero weights.
 900 Finally we analyze this 2 weight case and conclude it is bounded by $\frac{9}{8}$. Thus we have,

$$\sup_{t \geq 1} \sup_{y \in [0, 1]} \frac{\text{Var}[\hat{F}_{1:t}^{\text{comb}}]}{\text{Var}_{\text{opt}}(y, t)} \leq \frac{9}{8},$$

901

□

902 A.5 Proof of Proposition 5 (martingale CLT)

903 To construct the formal martingale CLT, we need to let the number of active measurement steps T
 904 and the domain size N go to infinity jointly. Suppose that for each N , we set the number of active
 905 measurement steps to be T_N ($T_N < N$) where T_N is non-decreasing with N . By letting $N \rightarrow \infty$ and
 906 $T_N \rightarrow \infty$, we have a martingale array $S_{N,t} = \sum_{\tau=1}^t X_{N,\tau} = \sum_{\tau=1}^t \tilde{\alpha}_\tau (\hat{F}_\tau - F(\Omega))$, for $t \leq T_N$.

907 Here $\tilde{\alpha}_\tau = \alpha_\tau / \sqrt{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2}$, $\text{Var}[\hat{F}_{\tau'}]$ is normalized using total variances. It is direct to check that
 908 $\mathbb{E}[S_{N,t+1} - S_{N,t} | S_{N,t}] = \mathbb{E}[X_{N,t+1} | S_{N,t}] = 0$. With addition assumptions, we are able to construct
 909 a limiting theorem for active measurement.

910 **Proposition 5** (formal). Assume that (1) $\sum_{\tau=1}^{T_N} \tilde{\alpha}_\tau^2 \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau] \rightarrow \eta^2$ as $N \rightarrow \infty$; (2) $P(\eta^2 <$
 911 $\infty) = 1$ and $P(\eta^2 > 0) = 1$; (3) $\sum_{\tau=1}^{T_N} \mathbb{E}[X_{N,\tau}^2 \mathbb{I}(|X_{N,\tau}| > \epsilon) | \mathcal{D}_\tau] \xrightarrow{p} 0$ for any $\epsilon > 0$. We have
 912 two central limit theorems

$$\frac{S_{N,T_N}}{\sqrt{\sum_{\tau=1}^{T_N} X_{N,\tau}^2}} \xrightarrow{D} \mathcal{N}(0, 1), \quad (11)$$

$$\frac{S_{N,T_N}}{\sqrt{\sum_{\tau=1}^{T_N} \tilde{\alpha}_\tau^2 \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]}} \xrightarrow{D} \mathcal{N}(0, 1). \quad (12)$$

913 *Proof.* We have met all the conditions for Theorem 3.3 and Corollary 3.2 in Hall and Heyde [13].
 914 Then we directly have

$$\begin{aligned} \frac{S_{N,T_N}}{U_{N,T_N}} & \xrightarrow{D} \mathcal{N}(0, 1), \\ \frac{S_{N,T_N}}{V_{N,T_N}} & \xrightarrow{D} \mathcal{N}(0, 1), \end{aligned}$$

915 where

$$\begin{aligned} U_{N,T_N} &= \sum_{\tau=1}^{T_N} X_{N,\tau}^2 \\ V_{N,T_N} &= \sum_{\tau=1}^{T_N} \tilde{\alpha}_\tau^2 \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]. \end{aligned}$$

917 The formal proposition employs a different normalizer from active measurement. It is useful for
 918 bounding the discrepancy between conditional and total variances. For active measurement, we have
 919 the following corollary.

920 **Corollary 1.** *With the same conditions as Prop. 5, we also have that*

$$\frac{\hat{F}_{1:T_N} - F(\Omega)}{\sqrt{\sum_{\tau=1}^{T_N} \bar{\alpha}_\tau^2 (\hat{F}_\tau - F(\Omega))^2}} \xrightarrow{D} \mathcal{N}(0, 1), \quad (13)$$

$$\frac{\hat{F}_{1:T_N} - F(\Omega)}{\sqrt{\sum_{\tau=1}^{T_N} \bar{\alpha}_\tau^2 \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]}} \xrightarrow{D} \mathcal{N}(0, 1). \quad (14)$$

921 *Proof.* Basically the equations are the same as Eq. 11 and 12, but with different normalizers. After
 922 multiplying $\frac{\sqrt{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \text{Var}[\hat{F}_{\tau'}]}}{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}}$ to both the numerator and denominator, we get Eq. 13 and 14. □

923 This leads to our informal version of the martingale CLT. Both $\text{Var}_{1:t}^{\text{cond}}$ and $\text{Var}_{1:t}^{\text{simp}}$ can be con-
 924 structed using the corollary.

925 We also demonstrate our assumptions in Prop. 5 with a simple example. If $\text{Var}[\hat{F}_\tau] = \sigma_\tau^2 =$
 926 $\lambda_1 \tau^{-y}/w_\tau$ and $\text{Var}[\hat{F}_\tau | \mathcal{D}_\tau] = \lambda_2 \tau^{-y}/w_\tau$ where $y \in [0, 1]$ and $0 < \lambda_2 \leq \lambda_1$, then with COMB
 927 weights

$$\begin{aligned} \sum_{\tau=1}^{T_N} \bar{\alpha}_\tau^2 \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau] &= \frac{\sum_{\tau=1}^{T_N} \lambda_2 w_\tau \tau^{1-y}}{\sum_{\tau=1}^{T_N} \lambda_1 w_\tau \tau^{1-y}} \\ &= \frac{\lambda_2}{\lambda_1}. \end{aligned}$$

928 It is clear that $0 < \lambda_2/\lambda_1 < \infty$, so the first two assumptions are met. For the third assumption, let
 929 $T_N = N/2$ and further assume for some $\delta > 0$, $\sup_{\tau, N} \frac{\mathbb{E}[(\hat{F}_\tau - F(\Omega))^{2+\delta} | \mathcal{D}_\tau]}{\sigma_\tau^{2+\delta}} < \infty$. The additional
 930 assumption implies the uniform boundedness of $\frac{(\hat{F}_\tau - F(\Omega))^2}{\sigma_\tau^2}$, which is common in the literature [38].
 931 Then note that

$$\begin{aligned} \frac{\max_\tau \alpha_\tau^2 \sigma_\tau^2}{\sum_{\tau=1}^{T_N} \alpha_\tau^2 \sigma_\tau^2} &= \frac{\lambda_1 w_{T_N} T_N^{1-y}}{\sum_{\tau=1}^{T_N} \lambda_1 w_\tau \tau^{1-y}} \\ &\leq \frac{\lambda_1 w_{T_N} T_N^{1-y}}{\frac{\lambda_1}{T_N} \left(\sum_{\tau=1}^{T_N} w_\tau \right) \left(\sum_{\tau=1}^{T_N} \tau^{1-y} \right)} \\ &= \frac{NT_N^{1-y}}{(N - T_N + 1) \left(\sum_{\tau=1}^{T_N} \tau^{1-y} \right)} \\ &\leq \frac{NT_N^{1-y}}{(N - T_N + 1) T_N^{2-y} / (2-y)} \rightarrow 0. \end{aligned} \quad (15)$$

932 Therefore, for any τ ,

$$\begin{aligned} \mathbb{E}[X_{N,\tau}^2 \mathbb{I}(|X_{N,\tau}| > \epsilon) | \mathcal{D}_\tau] &= \mathbb{E} \left[\frac{\alpha_\tau^2 (\hat{F}_\tau - F(\Omega))^2}{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \sigma_{\tau'}^2} \mathbb{I}(|X_{N,\tau'}| > \epsilon) | \mathcal{D}_\tau \right] \\ &= \frac{\alpha_\tau^2 \sigma_\tau^2}{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \sigma_{\tau'}^2} \mathbb{E} \left[\frac{(\hat{F}_\tau - F(\Omega))^2}{\sigma_\tau^2} \mathbb{I}(|X_{N,\tau'}| > \epsilon) | \mathcal{D}_\tau \right]. \end{aligned}$$

933 Note that

$$\mathbb{E} \left[\frac{(\hat{F}_\tau - F(\Omega))^2}{\sigma_\tau^2} \mathbb{I}(|X_{N,\tau'}| > \epsilon) | \mathcal{D}_\tau \right] = \mathbb{E} \left[\frac{(\hat{F}_\tau - F(\Omega))^2}{\sigma_\tau^2} \mathbb{I} \left(\frac{|\hat{F}_\tau - F(\Omega)|}{\sigma_\tau} > \epsilon \frac{\sqrt{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \sigma_{\tau'}^2}}{\alpha_\tau \sigma_\tau} \right) | \mathcal{D}_\tau \right].$$

934 Observe that $\frac{\mathbb{E}[(\hat{F}_\tau - F(\Omega))^2 | \mathcal{D}_\tau]}{\sigma_\tau^2} = \frac{\lambda_2}{\lambda_1}$, but $\frac{\sqrt{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \sigma_{\tau'}^2}}{\alpha_\tau \sigma_\tau} \rightarrow \infty$ for any τ according to Eq. 15. By
 935 the uniform boundedness, then for any $\epsilon_0 > 0$, there exists N_0 such that for any $N > N_0$,

$$\mathbb{E} \left[\frac{(\hat{F}_\tau - F(\Omega))^2}{\sigma_\tau^2} \mathbb{I}(|X_{N,\tau'}| > \epsilon) | \mathcal{D}_\tau \right] < \epsilon_0.$$

936 Therefore,

$$\begin{aligned} \sum_{\tau=1}^{T_N} \mathbb{E}[X_{N,\tau}^2 \mathbb{I}(|X_{N,\tau}| > \epsilon) | \mathcal{D}_\tau] &< \sum_{\tau=1}^{T_N} \frac{\alpha_\tau^2 \sigma_\tau^2}{\sum_{\tau'=1}^{T_N} \alpha_{\tau'}^2 \sigma_{\tau'}^2} \epsilon_0 \\ &= \epsilon_0. \end{aligned}$$

937 This shows that the third assumption is also met.

938 A.6 Proof of proposition 6

939 **Proposition 6.** *The estimators $\widehat{\text{Var}}_\tau$ and $\widehat{\text{Var}}_{\tau,r}$ for $\tau \leq r \leq t$ satisfy $\mathbb{E}[\widehat{\text{Var}}_\tau | \mathcal{D}_\tau] =$
 940 $\mathbb{E}[\widehat{\text{Var}}_{\tau,r} | \mathcal{D}_\tau] = \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]$ and $\mathbb{E}[\widehat{\text{Var}}_\tau] = \mathbb{E}[\widehat{\text{Var}}_{\tau,r}] = \text{Var}[\hat{F}_\tau]$.*

941 *Proof.* We first show the unbiasedness of individual variance estimators $\widehat{\text{Var}}_{\tau,r}$.

$$\begin{aligned} &\mathbb{E}[\widehat{\text{Var}}_{\tau,r} | \mathcal{D}_\tau] \\ &= \mathbb{E}_{\mathcal{D}_r \setminus \mathcal{D}_\tau, s_r} \left[\sum_{s \in \mathcal{D}_r \setminus \mathcal{D}_\tau} q_\tau(s) \left(\frac{f(s)}{q_\tau(s)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 + \frac{q_\tau(s_r)}{q_r(s_r)} \left(\frac{f(s_r)}{q_\tau(s_r)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 | \mathcal{D}_\tau \right] \\ &= \mathbb{E}_{\mathcal{D}_r \setminus \mathcal{D}_\tau} \left[\sum_{s \in \mathcal{D}_r \setminus \mathcal{D}_\tau} q_\tau(s) \left(\frac{f(s)}{q_\tau(s)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 + \mathbb{E}_{s_r} \left[\frac{q_\tau(s_r)}{q_r(s_r)} \left(\frac{f(s_r)}{q_\tau(s_r)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 \right] | \mathcal{D}_\tau \right] \\ &= \mathbb{E}_{\mathcal{D}_r \setminus \mathcal{D}_\tau} \left[\sum_{s \in \mathcal{D}_r \setminus \mathcal{D}_\tau} q_\tau(s) \left(\frac{f(s)}{q_\tau(s)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 + \sum_{s_r \in \Omega \setminus \mathcal{D}_\tau} q_\tau(s_r) \left(\frac{f(s_r)}{q_\tau(s_r)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 | \mathcal{D}_\tau \right] \\ &= \mathbb{E}_{\mathcal{D}_r \setminus \mathcal{D}_\tau} \left[\sum_{s \in \Omega \setminus \mathcal{D}_\tau} q_\tau(s) \left(\frac{f(s)}{q_\tau(s)} - F(\Omega \setminus \mathcal{D}_\tau) \right)^2 | \mathcal{D}_\tau \right] \\ &= \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]. \end{aligned}$$

942 Also, note that

$$\begin{aligned} \text{Var}_{\mathcal{D}_\tau} [\mathbb{E}[\hat{F}_\tau | \mathcal{D}_\tau]] &= \text{Var}_{\mathcal{D}_\tau} \left[\mathbb{E} \left[F(\mathcal{D}_\tau) + \frac{f(s_\tau)}{q(s_\tau)} | \mathcal{D}_\tau \right] \right] \\ &= \text{Var}_{\mathcal{D}_\tau} \left[\sum_{s_\tau \in \Omega \setminus \mathcal{D}_\tau} q(s_\tau) \left(F(\mathcal{D}_\tau) + \frac{f(s_\tau)}{q(s_\tau)} \right) \right] \\ &= \text{Var}_{\mathcal{D}_\tau} \left[F(\mathcal{D}_\tau) + \sum_{s_\tau \in \Omega \setminus \mathcal{D}_\tau} q(s_\tau) \frac{f(s_\tau)}{q(s_\tau)} \right] \\ &= \text{Var}_{\mathcal{D}_\tau} [F(\Omega)] \\ &= 0. \end{aligned}$$

943 Therefore, we further have

$$\begin{aligned}\mathbb{E}[\mathbb{E}[\widehat{\text{Var}}_{\tau,r} | \mathcal{D}_\tau]] &= \mathbb{E}_{\mathcal{D}_\tau}[\text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]] \\ &= \text{Var}[\hat{F}_\tau] - \text{Var}_{\mathcal{D}_\tau}[\mathbb{E}[\hat{F}_\tau | \mathcal{D}_\tau]] \\ &= \text{Var}[\hat{F}_\tau].\end{aligned}$$

944 Since $\widehat{\text{Var}}_\tau$ is convex on $\widehat{\text{Var}}_{\tau,r}$ with the weights, we also have $\mathbb{E}[\widehat{\text{Var}}_\tau | \mathcal{D}_\tau] = \text{Var}[\hat{F}_\tau | \mathcal{D}_\tau]$ and
945 $\mathbb{E}[\widehat{\text{Var}}_\tau] = \text{Var}[\hat{F}_\tau]$. \square

946 A.7 Proof of proposition 7

947 **Proposition 7.** *With the same settings as Prop. 3, the variance estimate for \hat{F}_τ weighted by*
948 *the LURE scheme $\widehat{\text{Var}}_\tau^{\text{LURE}} = \sum_{r=\tau}^t \bar{\beta}_r^{\text{LURE}} \widehat{\text{Var}}_{\tau,r}$ satisfies that $\text{Var}[\widehat{\text{Var}}_\tau^{\text{LURE}} | \mathcal{D}_\tau] \lesssim \frac{1}{t-\tau+1} \cdot$*
949 $\min\left(1, \frac{(N-t)^2}{t-\tau+1}\right)$.

950 *Proof.* We first derive the normalized LURE weights.

$$\begin{aligned}\sum_{r=\tau}^t \beta_r^{\text{LURE}} &= \sum_{r=\tau}^t \frac{1}{(N-r)(N-r+1)} \\ &= \sum_{r=\tau}^t \frac{1}{N-r} - \frac{1}{N-r+1} \\ &= \frac{1}{N-t} - \frac{1}{N-\tau+1} \\ &= \frac{t-\tau+1}{(N-t)(N-\tau+1)}.\end{aligned}$$

951 So the normalized LURE weights satisfy

$$\bar{\beta}_r^{\text{LURE}} = \frac{(N-t)(N-\tau+1)}{(t-\tau+1)(N-r)(N-r+1)}.$$

952 We then bound the variance of individual variance estimators. Observe that

$$\begin{aligned}\text{Var}[\widehat{\text{Var}}_{\tau,r} | \mathcal{D}_\tau] &= \text{Var}[\mathbb{E}[\widehat{\text{Var}}_{\tau,r} | D_r] | \mathcal{D}_\tau] + \mathbb{E}[\text{Var}[\widehat{\text{Var}}_{\tau,r} | D_r] | \mathcal{D}_\tau] \\ &= \text{Var}[\text{Var}[\hat{F}_\tau] | \mathcal{D}_\tau] + \mathbb{E}[\text{Var}[\widehat{\text{Var}}_{\tau,r} | D_r] | \mathcal{D}_\tau] \\ &= 0 + \mathbb{E}\left[\text{Var}\left[\frac{q_\tau(s_r)}{q_r(s_r)} \left(\frac{f(s_r)}{q_\tau(s_r)} - F(\Omega \setminus \mathcal{D}_\tau)\right)^2 \middle| D_r\right] \middle| \mathcal{D}_\tau\right] \\ &\leq \mathbb{E}\left[\mathbb{E}\left[\frac{q_\tau(s_r)^2}{q_r(s_r)^2} \left(\frac{f(s_r)}{q_\tau(s_r)} - F(\Omega \setminus \mathcal{D}_\tau)\right)^4 \middle| D_r\right] \middle| \mathcal{D}_\tau\right] \\ &\leq \mathbb{E}\left[\mathbb{E}\left[\frac{q_\tau(s_r)^2}{q_r(s_r)^2} \cdot 8 \cdot \left(\frac{f(s_r)^4}{q_\tau(s_r)^4} + F(\Omega \setminus \mathcal{D}_\tau)^4\right) \middle| D_r\right] \middle| \mathcal{D}_\tau\right] \\ &\leq \mathbb{E}\left[\mathbb{E}\left[\frac{B^4(N-r+1)^2}{A^4(N-\tau+1)^2} \cdot 8 \cdot \left(\frac{B^4(N-\tau+1)^4 B^4}{A^4} + B^4(N-\tau+1)^4\right) \middle| D_r\right] \middle| \mathcal{D}_\tau\right] \\ &= \frac{8B^8(B^4 + A^4)}{A^8} (N-r+1)^2 (N-\tau+1)^2.\end{aligned}$$

953 Therefore, let $D = \frac{8B^8(B^4 + A^4)}{A^8}$, we can say that $\text{Var}[\widehat{\text{Var}}_{\tau,r} | \mathcal{D}_\tau] \leq D(N-\tau+1)^2(N-r+1)^2$. Next,
954 we show that for every two variance estimators $\text{Cov}(\widehat{\text{Var}}_{\tau,r_1}, \widehat{\text{Var}}_{\tau,r_2} | \mathcal{D}_\tau) = 0$ if $\tau \leq r_1 < r_2 \leq N$.

$$\begin{aligned}\text{Cov}(\widehat{\text{Var}}_{\tau,r_1}, \widehat{\text{Var}}_{\tau,r_2} | \mathcal{D}_\tau) &= \mathbb{E}[(\widehat{\text{Var}}_{\tau,r_1} - \text{Var}_{s_\tau}[\hat{F}_\tau | \mathcal{D}_\tau])(\widehat{\text{Var}}_{\tau,r_2} - \text{Var}_{s_\tau}[\hat{F}_\tau | \mathcal{D}_\tau]) | \mathcal{D}_\tau] \\ &= \mathbb{E}_{\mathcal{D}_{r_2} \setminus \mathcal{D}_\tau}[(\widehat{\text{Var}}_{\tau,r_1} - \text{Var}_{s_\tau}[\hat{F}_\tau | \mathcal{D}_\tau]) \mathbb{E}_{s_{r_2}}[(\widehat{\text{Var}}_{\tau,r_2} - \text{Var}_{s_\tau}[\hat{F}_\tau | \mathcal{D}_\tau])] \\ &= \mathbb{E}_{\mathcal{D}_{r_2} \setminus \mathcal{D}_\tau}[(\widehat{\text{Var}}_{\tau,r_1} - \text{Var}_{s_\tau}[\hat{F}_\tau | \mathcal{D}_\tau]) \cdot 0] \\ &= 0.\end{aligned}$$

955 Therefore (when $t < N$),

$$\begin{aligned}\text{Var} [\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] &= \sum_{r=\tau}^t \frac{(N-t)^2(N-\tau+1)^2}{(t-\tau+1)^2(N-r)^2(N-r+1)^2} \text{Var}[\widehat{\text{Var}}_{\tau,r}] \\ &\leq \sum_{r=\tau}^t \frac{(N-t)^2(N-\tau+1)^2}{(t-\tau+1)^2(N-r)^2(N-r+1)^2} D(N-\tau+1)^2(N-r+1)^2 \\ &= \frac{D(N-t)^2(N-\tau+1)^4}{(t-\tau+1)^2} \sum_{r=\tau}^t \frac{1}{(N-r)^2}.\end{aligned}$$

956 There are two ways to bound the right hand side. When t is far from N , we can bound with

$$\begin{aligned}\text{Var} [\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] &\leq \frac{D(N-t)^2(N-\tau+1)^4}{(t-\tau+1)^2} \sum_{r=\tau}^t \frac{1}{(N-r)^2} \\ &\leq \frac{D(N-t)^2(N-\tau+1)^4}{(t-\tau+1)^2} \sum_{r=\tau}^t \frac{1}{(N-t)^2} \\ &= \frac{D(N-\tau+1)^4}{t-\tau+1}.\end{aligned}$$

957 When t is close to N , by the fact that $\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}$, we have

$$\begin{aligned}\text{Var} [\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] &\leq \frac{D(N-t)^2(N-\tau+1)^4}{(t-\tau+1)^2} \sum_{r=\tau}^t \frac{1}{(N-r)^2} \\ &\leq \frac{D(N-\tau+1)^4}{t-\tau+1} \cdot \frac{\pi^2(N-t)^2}{6(t-\tau+1)}.\end{aligned}$$

958 The above two inequalities still hold when $t = N$, where LURE will assign infinite weights
959 on the exact variance $\widehat{\text{Var}}_{\tau,N} = \text{Var}[\hat{F}_{\tau}|D_{\tau}]$, so $\text{Var}[\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] = 0$. In summary, we
960 see that $\text{Var} [\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] \leq \frac{E}{t-\tau+1} \cdot \min\left(1, \frac{(N-t)^2}{(t-\tau+1)}\right)$, for $E = \frac{\pi^2 D(N-\tau+1)^2}{6}$. Thus
961 $\text{Var} [\widehat{\text{Var}}_{\tau}^{\text{LURE}} | \mathcal{D}_{\tau}] \lesssim \frac{1}{t-\tau+1} \cdot \min\left(1, \frac{(N-t)^2}{(t-\tau+1)}\right)$.

962 □

963 B Details of variance estimation

964 B.1 The plug-in mean in variance estimation

965 One practical gap is the incorporation of a plug-in mean instead of the ground-truth mean in Eq. 6.
966 We show that the error introduced by this trick is controlled. First of all, the variance of the plug-in
967 mean could be controlled with the following lemma.

968 **Lemma 1.** For $\hat{F}_{1:t}^{\text{LURE}} = \sum_{\tau=1}^t \bar{\alpha}_{\tau}^{\text{LURE}} \hat{F}_{\tau}$, its variance satisfies that there exists a constant $E > 0$,
969 $\text{Var}[\hat{F}_{1:t}^{\text{LURE}}] \leq \frac{E(N-t)}{t}$.

970 *Proof.* By Prop. 3,

$$\text{Var}[\hat{F}_{\tau}] \leq C(N-\tau)(N-\tau+1).$$

971 Then

$$\begin{aligned}
\text{Var}[\hat{F}_{1:t}^{\text{LURE}}] &= \sum_{\tau=1}^t (\bar{\alpha}_{\tau}^{\text{LURE}})^2 \text{Var}[\hat{F}_{\tau}] \\
&\leq \sum_{\tau=1}^t \frac{N^2(N-t)^2}{t^2(N-\tau)^2(N-\tau+1)^2} \cdot C(N-\tau)(N-\tau+1) \\
&= \frac{CN^2(N-t)^2}{t^2} \sum_{\tau=1}^t \frac{1}{(N-\tau)(N-\tau+1)} \\
&= \frac{CN^2(N-t)^2}{t^2} \frac{t}{N(N-t)} \\
&= \frac{CN(N-t)}{t}.
\end{aligned}$$

972 Let $E = CN$, then $\text{Var}[\hat{F}_{1:t}^{\text{LURE}}] \leq \frac{E(N-t)}{t}$. □

973 This directly implies that the plug-in mean constructed from the LURE estimate will not be far from
974 $F(\Omega)$. Denote the variance estimate with the plug-in mean by

$$\widetilde{\text{Var}}_{\tau,t} = \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_{\tau}} q_{\tau}(s) \left(\frac{f(s)}{q_{\tau}(s)} - \hat{G}_{\tau,t} \right)^2 + \frac{q_{\tau}(s_t)}{q_t(s_t)} \left(\frac{f(s_t)}{q_{\tau}(s_t)} - \hat{G}_{\tau,t} \right)^2.$$

975 We control the error with the following proposition.

976 **Proposition 8.** *Given the same assumptions as Prop. 3, with at least probability $1 - p$, there exists a*
977 *constant H dependent on p such that $|\widetilde{\text{Var}}_{\tau,t} - \widehat{\text{Var}}_{\tau,t}| \leq H(N - \tau + 1) \sqrt{\frac{N-t+1}{t-1}}$.*

978 *Proof.* First, note that

$$\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_{\tau}) = \hat{F}_{1:t}^{\text{LURE}} - F(\Omega).$$

979 By Chebyshev's inequality, with at least probability $1 - p$,

$$|\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_{\tau})| = |\hat{F}_{1:t-1}^{\text{LURE}} - F(\Omega)| \leq \sqrt{\frac{E(N-t+1)}{p(t-1)}}.$$

980 Also,

$$\begin{aligned}
|\hat{G}_{\tau,t} + F(\Omega \setminus \mathcal{D}_{\tau})| &\leq |\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_{\tau})| + 2F(\Omega \setminus \mathcal{D}_{\tau}) \\
&\leq \sqrt{\frac{E(N-t+1)}{p(t-1)}} + 2B(N - \tau + 1) \\
&\leq I(N - \tau + 1),
\end{aligned}$$

981 where I is a constant dependent on p . We expand the difference directly.

$$\begin{aligned}
|\widehat{\text{Var}}_{\tau,t} - \widehat{\text{Var}}_{\tau,t}| &= \left| \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} q_\tau(s) (\hat{G}_{\tau,t}^2 - F(\Omega \setminus \mathcal{D}_\tau)^2) - 2f(s) (\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_\tau)) \right. \\
&\quad \left. + \frac{q_\tau(s_t)}{q_t(s_t)} \left(\hat{G}_{\tau,t}^2 - F(\Omega \setminus \mathcal{D}_\tau)^2 \right) - 2 \frac{f(s_t)}{q_t(s_t)} (\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_\tau)) \right| \\
&\leq \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} q_\tau(s) |\hat{G}_{\tau,t}^2 - F(\Omega \setminus \mathcal{D}_\tau)^2| + 2f(s) |\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_\tau)| \\
&\quad + \frac{q_\tau(s_t)}{q_t(s_t)} |\hat{G}_{\tau,t}^2 - F(\Omega \setminus \mathcal{D}_\tau)^2| + 2 \frac{f(s_t)}{q_t(s_t)} |\hat{G}_{\tau,t} - F(\Omega \setminus \mathcal{D}_\tau)| \\
&\leq \sqrt{\frac{E(N-t+1)}{p(t-1)}} \left(\sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} \frac{B}{(N-\tau+1)A} I(N-\tau+1) + 2B \right. \\
&\quad \left. + \frac{B^2(N-t+1)}{A^2(N-\tau+1)} I(N-\tau+1) + 2 \frac{B^2(N-t+1)}{A} \right) \\
&= \left(\frac{IB}{A} + 2B \right) \left(t - \tau + \frac{B}{A}(N-t+1) \right) \sqrt{\frac{E(N-t+1)}{p(t-1)}} \\
&\leq \frac{(I+2A)B^2}{A^2} \sqrt{\frac{E}{p}} (N-\tau+1) \sqrt{\frac{N-t+1}{t-1}}.
\end{aligned}$$

982 Let $H = \frac{(I+2A)B^2}{A^2} \sqrt{\frac{E}{p}}$, then $|\widehat{\text{Var}}_{\tau,t} - \widehat{\text{Var}}_{\tau,t}| \leq H(N-\tau+1) \sqrt{\frac{N-t+1}{t-1}}$. \square

983 Comparing the derivation of Prop. 7 and Prop. 8, the L1 error introduced by the plug-in mean goes to
984 0 at a faster rate than the standard deviation of $\widehat{\text{Var}}_{\tau,t}$, indicating that the error is negligible in our
985 evaluation when t is large enough.

986 B.2 The streaming algorithm of variance estimation

987 Alg. 2 has a summation for each $1 \leq \tau \leq t$, amounting to a complexity of $\mathcal{O}(t^2)$. In practice, we
988 would reuse part of previous computation at step $t-1$ to accelerate the computation at step t . Observe
989 that

$$\begin{aligned}
\widehat{\text{Var}}_{\tau,t} &= \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} q_\tau(s) \left(\frac{f(s)}{q_\tau(s)} - \hat{G}_{\tau,t} \right)^2 + \frac{q_\tau(s_t)}{q_t(s_t)} \left(\frac{f(s_t)}{q_\tau(s_t)} - \hat{G}_{\tau,t} \right)^2 \\
&= \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} \left(\frac{f(s)^2}{q_\tau(s)} + q_\tau(s) \hat{G}_{\tau,t}^2 - 2f(s) \hat{G}_{\tau,t} \right) + \frac{q_\tau(s_t)}{q_t(s_t)} \left(\frac{f(s_t)}{q_\tau(s_t)} - \hat{G}_{\tau,t} \right)^2 \\
&= \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} \frac{f(s)^2}{q_\tau(s)} + \hat{G}_{\tau,t}^2 \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} q_\tau(s) - 2\hat{G}_{\tau,t} \sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} f(s) + \frac{q_\tau(s_t)}{q_t(s_t)} \left(\frac{f(s_t)}{q_\tau(s_t)} - \hat{G}_{\tau,t} \right)^2.
\end{aligned}$$

990 This is a polynomial with respect to our plug-in mean $\hat{G}_{\tau,t}$. Therefore, we could maintain the three
991 summations in the coefficients $\sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} \frac{f(s)^2}{q_\tau(s)}$, $\sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} q_\tau(s)$ and $\sum_{s \in \mathcal{D}_t \setminus \mathcal{D}_\tau} f(s)$ for each τ
992 during the algorithm to speed-up the computation of each individual $\widehat{\text{Var}}_{\tau,t}$. In addition, our final
993 estimate is

$$\widehat{\text{Var}}_\tau = \sum_{r=\tau}^t \bar{\beta}_r \widehat{\text{Var}}_{\tau,r} = \frac{\sum_{r=\tau}^t \beta_r \widehat{\text{Var}}_{\tau,r}}{\sum_{r=\tau}^t \beta_r}.$$

994 The numerator is still a polynomial with respect to the plug-in mean whose coefficients can be written
995 as summations. We would further maintain the three coefficients and $\sum_{r=\tau}^t \beta_r$ to avoid the loop to
996 compute $\widehat{\text{Var}}_\tau$. With the two techniques, our approach is summarized as Alg. 3.

Algorithm 3 Streaming variance estimation update

Require: Sample s_t , label $f(s_t)$ and estimate $\hat{F}_{1:t-1}$

```
1: Set  $x_t = y_t = z_t = a_t = b_t = c_t = u_t = 0$ 
2: for  $\tau = 1, 2, \dots, t$  do
3:   Get estimated mean  $\hat{G}_{\tau,t} = \hat{F}_{1:t-1} - F(\mathcal{D}_\tau)$ 
4:    $a_\tau \leftarrow a_\tau + \beta_t \left( x_\tau + \frac{f(s_t)^2}{q_t(s_t)q_\tau(s_t)} \right)$ 
5:    $b_\tau \leftarrow b_\tau + \beta_t \left( y_\tau + \frac{f(s_t)}{q_t(s_t)} \right)$ 
6:    $c_\tau \leftarrow c_\tau + \beta_t \left( z_\tau + \frac{q_\tau(s_t)}{q_t(s_t)} \right)$ 
7:    $u_\tau \leftarrow u_\tau + \beta_t$ 
8:    $\widehat{\text{Var}}_\tau = (a_\tau - 2b_\tau \hat{G}_{\tau,t} + c_\tau \hat{G}_{\tau,t}^2) / u_\tau$ 
9:    $x_\tau \leftarrow x_\tau + \frac{f(s_t)^2}{q_\tau(s_t)}$ 
10:   $y_\tau \leftarrow y_\tau + f(s_t)$ 
11:   $z_\tau \leftarrow z_\tau + q_\tau(s_t)$ 
12: end for
```

997 C Experimental settings

998 C.1 Estimating birds in radar data

999 **Weather radar preliminaries.** The NEXRAD radar network contains 159 high-resolution radars
1000 and covers most of the U.S. territories. Each radar station typically scans the surrounding atmosphere
1001 every 4-10 minutes to collect weather data. This is achieved by rotating the radar antenna around the
1002 vertical axis at multiple elevation angles. At each elevation, the radar performs “sweeps” to collect
1003 several radar products, such as, reflectivity and radial velocity. These signals also record objects like
1004 bird roosts. We render radar data for $300\text{km} \times 300\text{km}$ regions into 600×600 pixel arrays, on which
1005 we train a detector model to automatically predict bounding boxes for roosts.

1006 **Visualization.** Fig. A2 (left) illustrates weather radar scans capturing bird roosts. Fig. A2 (right)
1007 visualizes the reflectivity data at 0.5° elevation of a radar scan that happened at the KCLE weather
1008 radar station on August 19 2015 at 10:31:22 UTC, which is a randomly sampled radar scan that
1009 captures bird roosts. The circular and semicircular patterns are massive amounts of birds departing
1010 from their overnight roosting locations.

1011 **Detector, tracking, and filtering.** We follow Perez et al. [26] to pretrain a station-agnostic spa-
1012 tiotemporal roost detector. We also follow their configuration for deploying the detector on the Great
1013 Lakes radar stations, which are excluded from the pretraining data, assembling model predicted
1014 bounding boxes into roost tracks, and filtering tracks with too few detected time frames or too low of
1015 a max or mean detection confidence score. Since each weather radar collects data every few minutes,
1016 the same roost is often captured by multiple consecutive scans of a radar station. To avoid double
1017 counting, it is necessary to assemble detections of the same roost into a track.

1018 Perez et al. [26] deploys the roost prediction system on radar data from 12 radar stations in the Great
1019 Lakes region and let human experts screen the system predictions. They focus on the time window
1020 from 30 minutes to 90 minutes after the local sunrise for every station-day.

1021 Following Deng et al. [8], we focus on 11 stations, including KAPX, KBUF, KCLE, KDLH, KDTX,
1022 KGRB, KGRR, KLOT, KMKX, KTYX, KIWX, since KMQT does not clearly observe roosts in the
1023 human screened data. Our goal is to estimate the number of birds in June to October in the 2015-2019
1024 five-year period at these 11 stations. We use screened data from Perez et al. [26] as the ground truth
1025 bounding boxes for estimating bird counts. We let our station-specific finetuned checkpoints predict
1026 bounding boxes for the same time periods on the same station-day for our estimations.

1027 Station-day bird count estimation.

- 1028 • Per-sweep counts. For each predicted bounding box by the detector model in a radar scan in
1029 a station-day, we enumerate over the radar sweeps at all elevation within 5000km height and

- 1030 follow Belotti et al. [4] to estimate the bird count of this bounding box geographical region
 1031 in this sweep.
- 1032 • Per-bounding-box counts. We summarize, for each detection, radar sweeps taken across
 1033 multiple elevations. In order to prevent double counting of birds in regions sampled twice by
 1034 two consecutive beams, we bin the sweep elevations into 10 bins. We then take the average
 1035 of sweeps that fall in the same bin. Lastly, we sum counts across all bins.
 - 1036 • Per-track counts. We obtain summaries across detections by finding the median count within
 1037 each track and selecting two scans before and after the median. We then calculate the
 1038 average count within this set of 5 scans.
 - 1039 • Per-day counts. We sum over per-track counts to obtain a per-day bird count at a station-day.

1040 C.2 Additional Finetuning And Sampling Details

1041 To fine-tune our models, we use the Detectron2 library with modified configurations. Specifically, we
 1042 disable learning rate warmup, set the batch size to 8, and set `FILTER_EMPTY_ANNOTATIONS`
 1043 to `False`. For the high-resolution image experiments, we use the `faster_rcnn_R_50_FPN_3x.yaml`
 1044 configuration with the default ImageNet-pretrained weights.

1045 Fine-tuning the image detector takes approximately 20 minutes on a single NVIDIA A16 GPU. We
 1046 perform fine-tuning 8 times, resulting in a total end-to-end runtime of just under 3 hours per image.
 1047 For the roost detector, each fine-tuning run takes about 2 hours on a single A16 GPU. We fine-tune
 1048 4 times, totaling roughly 8 hours of compute per station for roost counting. In both cases, the time
 1049 required to compute count and variance estimates is negligible.

1050 In practice, the measurements from the detectors could be as low as 0, making the labeling of some
 1051 units impossible and introducing huge variances to the estimation. In our image counting tasks, we
 1052 set $g(s) \leftarrow \max(g(s), 1)$ to make sure that each tile can be sampled. In our radar counting tasks, we
 1053 set $g(s) \rightarrow g(s) + 1000$ to cover every day in every station.

1054 C.3 Labeling Birds in Images

1055 We visualize the Reeds and Sky images in figure A1. To collect ground truth, we manually label both
 1056 images entirely. We use a tile size of 200 pixels for sky and 160 pixels for reeds, resulting in a total
 1057 of 925 and 1426 tiles respectively. To label we randomly group the tiles into batches of 50, and have
 1058 labelers annotate each batch using VGG annotator. To speed up the process, we label each bird with
 1059 a single point at its center of mass. This results in about 1 second of human effort per labeled point.
 1060 We only label a bird on the edge if a majority of it is visible in the tile. In total, our ground truth
 1061 count for sky is 5682 birds and 12486 birds for reeds.

1062 To convert point annotations into approximate bounding boxes, we use a two-stage heuristic algorithm.
 1063 First, we apply Otsu thresholding to each tile and compute the average size of a square bounding
 1064 box by dividing the number of foreground pixels by the number of labeled points. In the second
 1065 stage, we center these boxes at each labeled point, mask out foreground pixels outside the boxes, and
 1066 recompute the average box size using the remaining foreground pixels.

1067 D Additional experimental results

1068 D.1 Comparison to few-shot models

1069 Figure A3 presents results in terms of the Hellinger distance between the predicted proposal dis-
 1070 tribution P and the optimal sampling distribution Q (proportional to ground truth counts). We use
 1071 Hellinger distance because it accommodates zero-probability entries. It is defined as:

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}.$$

1072 Where p_i, q_i are the probabilities assigned to unit i for P and Q , respectively. Unlike raw counts, the
 1073 Hellinger distance is unaffected by constant-factor over or under prediction and thus gives a better
 1074 idea of how good the model is for the estimation process.



Figure A1: The Sky image (left) and Reeds image (right) used for our detection experiments. Note the resolution has been reduced for this visualization.

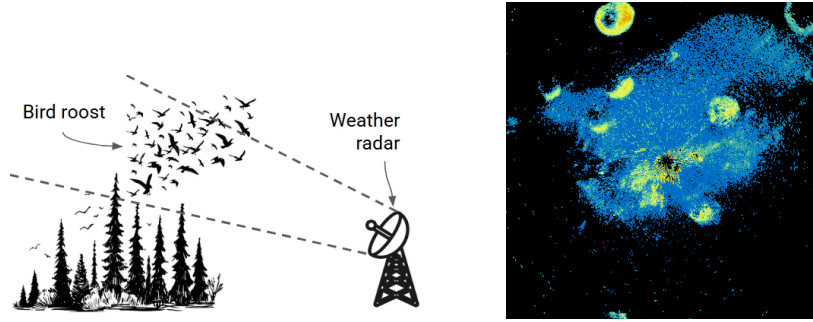


Figure A2: (Left) Weather radars can detect bird roosts departing from their nighttime roosting locations. (Right) Roosts appear as circular patterns in images rendered from radar products, such as the reflectivity at 0.5° channel as visualized.

1075 On the left, we compare the proposal distributions of the sky image raw predictions from our fine-
 1076 tuned detector and the few-shot FamNet model [29], which estimates object counts from a few
 1077 examples without additional training. FamNet struggles to adapt with more examples, while the
 1078 fine-tuned model achieves lower Hellinger distance even with a single labeled tile.

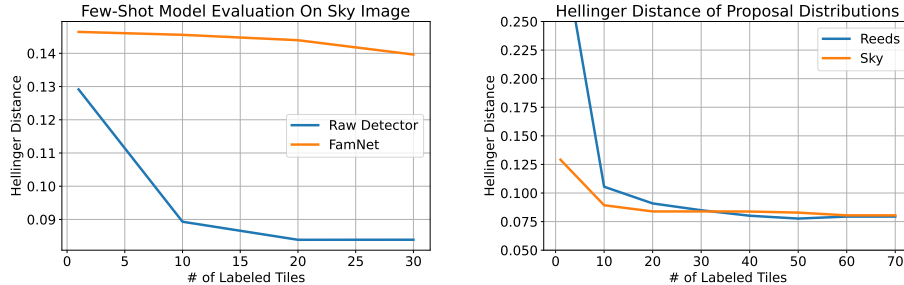


Figure A3: Left: Comparing the performance of raw predictions from fine-tuned detectors vs. few-shot FamNet[29] model, based on Hellinger distance to ground truth. The FamNet model performs worse and does not improve as much with more labeled tiles. Right: Hellinger distance between the raw predictions of the fine-tuned detector and the optimal sampling distribution for both images, averaged over 100 checkpoints. Unlike the raw counts which saturate quickly, additional training continues to improve the proposal distribution, especially on the harder Reeds image.

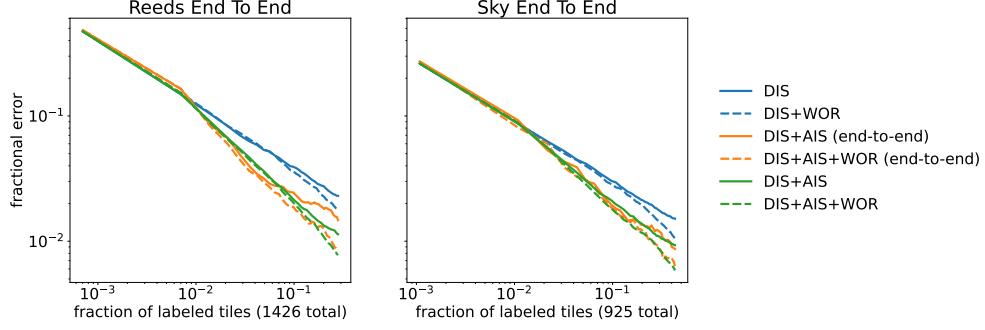


Figure A4: Comparing our fixed checkpoint scheme against true end-to-end training, averaged over 1000 trials. While the fixed checkpoint approach introduces slightly higher variance, particularly at higher label counts, it preserves relative trends between methods. This supports its use for our large-scale evaluation.

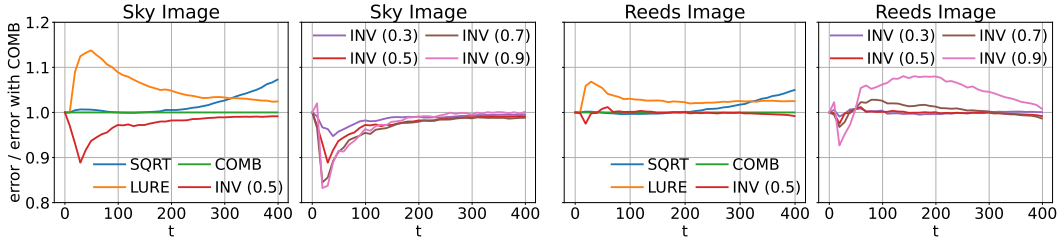


Figure A5: Relative errors compared to α^{COMB} weighting. Other fixed weighting strategies (α^{SQRT} , α^{LURE}) are worse, but inverse variance weighting (denoted by $\text{INV}(\gamma)$) may achieve lower error.

D.2 Fine-Tuning Improves Proposal Distribution

On the right of Figure A3, we track the Hellinger distance over the course of finetuning, averaged over 100 checkpoints. While raw detector counts tend to saturate early, the Hellinger distance of the proposal distribution continues to improve, especially for the harder Reeds image, before eventually saturating.

D.3 True End-to-End Finetuning

To approximate the effect of interactive model adaptation in our main results, we use a “fixed checkpoint” approach: a predefined sequence of detectors trained on a progressively larger number of labels. This allows us to conduct a significantly larger number of trials than would be feasible with full end-to-end training.

In practice, however, a practitioner would retrain the detector as new samples are labeled during active measurement. To assess how well our “fixed checkpoint” approach approximates this more realistic scenario, we compare it to end-to-end training averaged over 1,000 trials (Figure A4).

The results indicate that performance under both approaches is similar, particularly when using fewer tiles, and that the relative trends between methods are preserved. We do observe slightly higher variance in the fixed checkpoint setting, likely due to the specific checkpoint chosen.

D.4 Different weighting schemes for all experiments

We compare different weighting schemes in both image counting and radar counting experiments. The results are in Fig. A5 and A6. From the first and the third columns, we conclude that the α^{COMB} weights work consistently better than α^{SQRT} and α^{LURE} . In the second and the fourth columns, the performances of inverse variance weighting with different γ are compared. In general, a conservative $\gamma = 0.3$ brings little benefit, while an aggressive $\gamma = 0.9$ can be detrimental because each individual estimator may not be accurate enough.

1102 D.5 Confidence intervals for all experiments

1103 In Fig. A7, we compare the CI coverages for all experiments. In most of the experiments, the
1104 coverage for active measurement improves as the number of labeled samples increases (except on
1105 KIWX and KTYX). This behavior is typical in importance sampling-based methods. Out of the 11
1106 radar stations, we achieve near perfect coverage (around 0.95) in 6, good coverage (around 0.9) in 2
1107 and fair coverage (around 0.8) in 3. Comparing the two variance estimators $\text{Var}_{1:t}^{\text{simp}}$ and $\text{Var}_{1:t}^{\text{cond}}$, we
1108 find that the $\text{Var}_{1:t}^{\text{simp}}$ works better, especially on stations with bad coverage. We conclude that the
1109 CIs in active measurement usually have the desired properties for measuring the uncertainty.

1110 D.6 More comparisons with other baselines

1111 We also compare active measurement with two other baselines: active testing and \hat{H}_t (motivated by
1112 PPI), on the image counting tasks. The results are in Fig. A8. In general, we find active measurement
1113 has the lowest error.

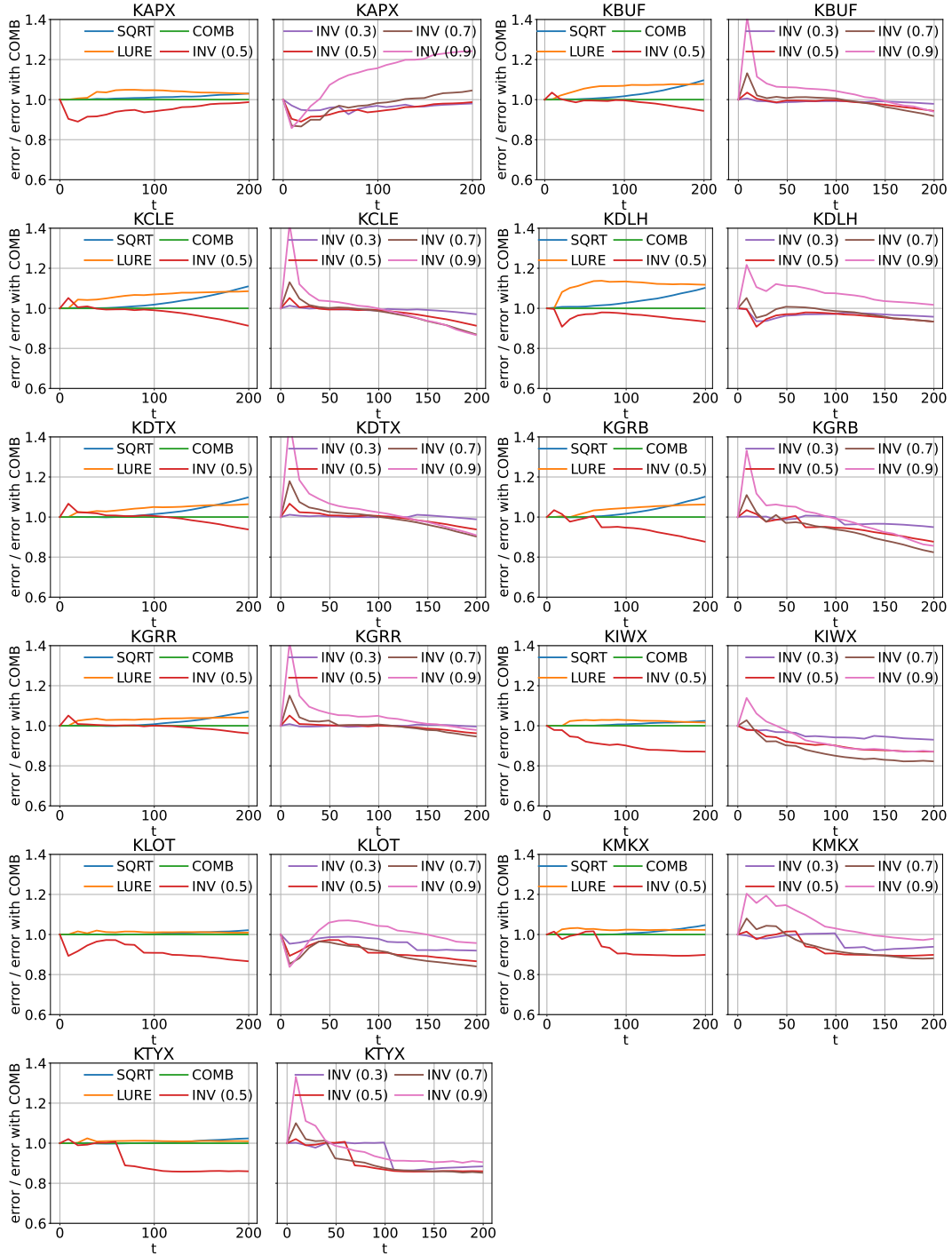


Figure A6: Relative errors compared to α^{COMB} weighting, for the roost counting problem on all 11 radar stations.

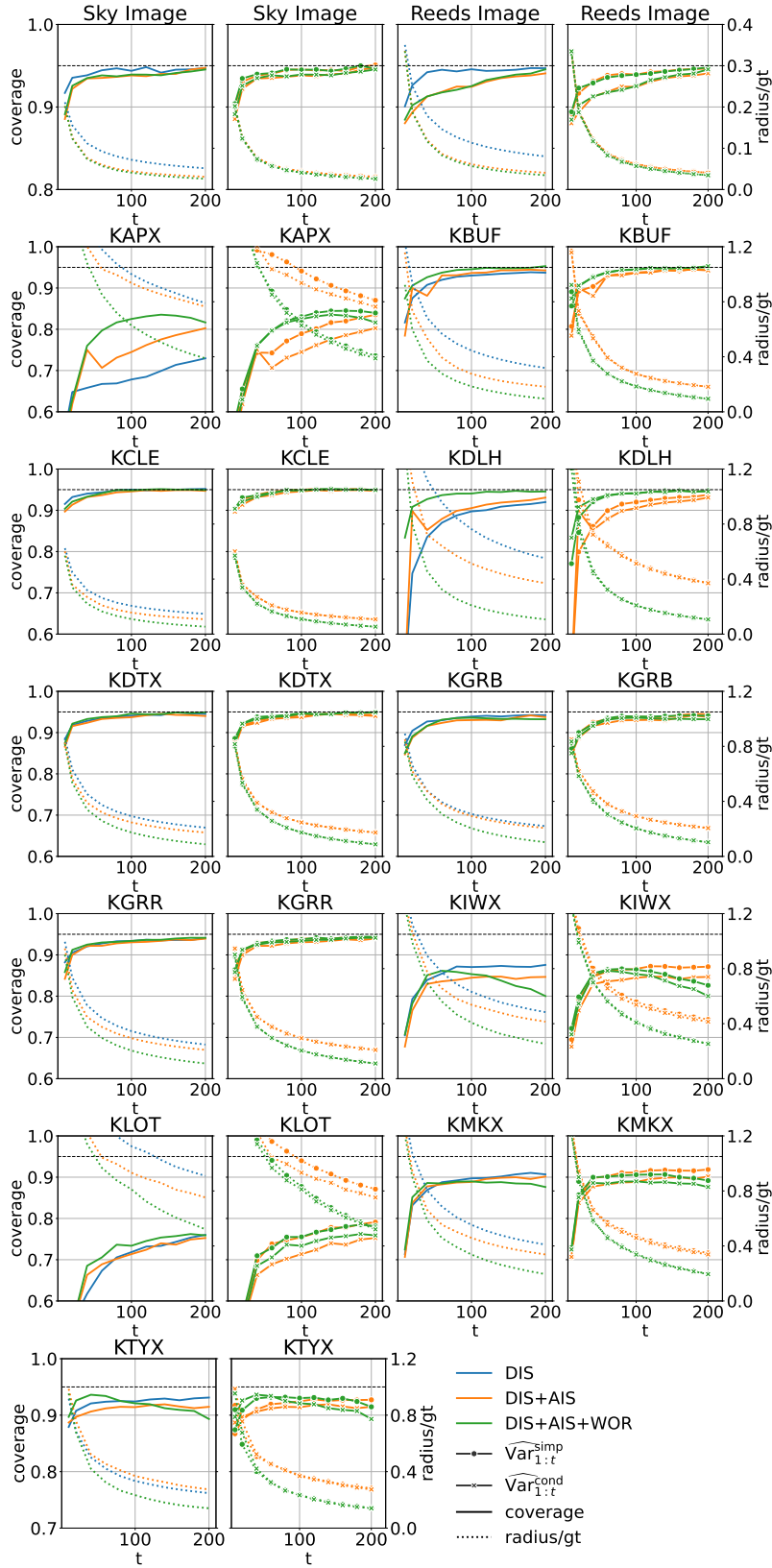


Figure A7: Full CI coverage results for counting on the two images and on all 11 radar stations.

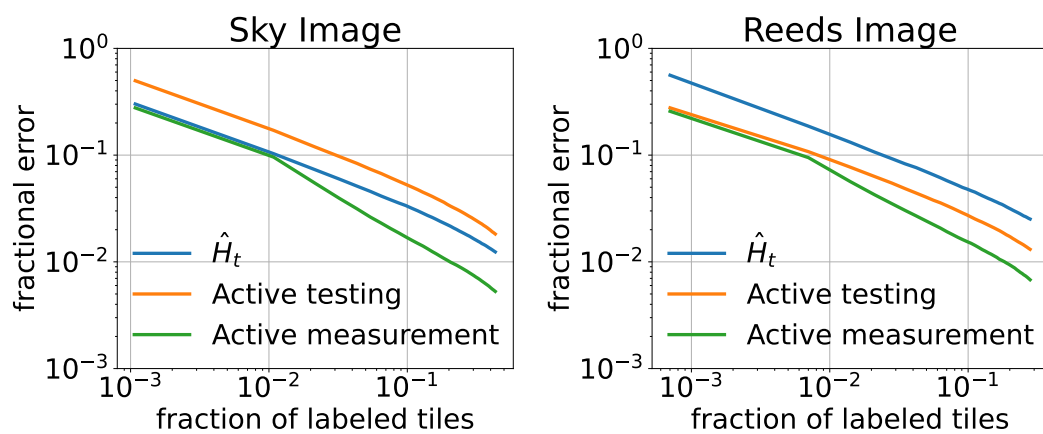


Figure A8: Fractional error compared with other baselines, for the two image counting tasks, averaged over 10,000 simulations.