RECALIBRATED ROUTING FOR LLM JUDGES: OPTIMIZING CLASSIFICATION METRICS ON A BUDGET

Anonymous authors
Paper under double-blind review

ABSTRACT

Routing is important in LLM-as-a-judge research because it combines models with different comparative strengths in a cost-effective manner. In practice, performance of such routers is often measured using interpretable classfication metrics such as precision, recall, and F_1 score; however, existing methods typically optimize objectives in the form of average proxy quality that need not align with these deployment metrics. In this paper, we develop ReCal-Route, a budget-aware framework that directly optimizes the chosen metric. In particular, we use a quasi-linear representation to formulate routing as a mixed-integer linear program with nuisance parameters. Furthermore, post-call recalibration uses the returned prediction to improve decisions without additional model calls. For theoretical guarantees, we derive regret bounds in terms of probability and cutoff estimation error. We apply the method on a real-world safety violation detection problem at a tech company and show that ReCal-Route outperforms existing methods in the literature.

1 INTRODUCTION

LLM-as-a-judge research has emerged as an important topic in various areas, such as content moderation, due to the large volume of items to evaluate at a high velocity. As AI tools become more diverse, choosing cost-effective LLM judges becomes increasingly important due to the vertical and horizontal differentiation. On the vertical front, calling an expensive LLM for every item may exhaust the budget, but a cheaper judge may miss violations that another LLM would catch. On the horizontal front, different LLM judges may align more closely with human assessment in different domains and on different items. Selecting LLM judges in a cost-effective fashion can be cast as a routing problem: given the information available before any LLM is called, which judge should assess the item, and should a second judge be consulted? The goal is to allocate calls across items to improve classification performance within a given budget.

While the routing perspective is not new (e.g., Chen et al., 2024; Hu et al., 2024; Ong et al., 2025; Dekoninck et al., 2025; Zhang et al., 2025; Li et al., 2026), we identify two major limitations. First, existing methods largely optimize average per-query quality, for example by minimizing 0–1 loss or maximizing a proxy quality score. These objectives need not reflect the different consequences of false positives and false negatives, and need not align with classification metrics such as precision, recall, and F_1 that are often used in practice. In particular, precision and F_1 are nonlinear functions of aggregate classification rates and generally cannot be written as averages or expectations of fixed per-query scores (e.g., Koyejo et al., 2014; Puthiya Parambath et al., 2014; Kotłowski and Dembczyński, 2016). This misalignment matters for safety screening in content moderation, where most content is safe: high accuracy can coexist with a high false-negative rate and poor detection of violations. Conversely, excessive false positives may discourage content producers and reduce the production of legitimate content. We therefore optimize the classification metric directly, with F_β being our primary focus.

Second, using the selected judge's raw label can discard information returned by the LLM. For example, it is common to ask LLMs to return a confidence score along with the answer. Confidence scores need not be calibrated probabilities (Guo et al., 2017), but they may still contain information about the true label. In addition, pre-call context, such as content type or source, may be associated with how conservative an LLM's judgments are. Instead of using the selected judge's raw label as

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the final decision, we add a post-routing layer that recalibrates the LLM judgment by training a simple prediction model for the ground-truth label using the LLM label, pre-call context, confidence scores, and any other auxiliary information available after the LLM call. Its decision class includes the common practical rule of using only the LLM's raw label. Thus, at the population optimum, recalibration never hurts the final performance and can strictly improve upon the uncalibrated version when the raw labels yield suboptimal decisions for the deployment metric. This is analogous to the no-harm calibration in the causal inference literature (Lei and Ding, 2018; Cohen and Fogarty, 2024).

Contributions. **(i) Framework:** We formulate recalibrated routing for LLM judges on a budget as a constrained nonlinear policy learning problem (Athey and Wager, 2021). We use classification metrics as objectives and focus on F_β, a generalization of the F_1 score. **(ii) Method:** We propose ReCal-Route, a two-stage procedure using a static dataset of item contexts, ground-truth labels, and evaluations from a pool of candidate LLM judges. In the first stage, we cross-fit the conditional probability of a positive label given the context and the outputs available for each LLM judge candidate, including raw labels and confidence scores. In the second stage, we use these nuisance estimates to construct recalibrated decisions and estimate the F_β score of candidate routing policies. We use the linear-fractional structure of F_β to formulate successive routing updates as mixed-integer linear programs (MILPs). **(iii) Theory:** We establish a vanishing regret bound for the routing performance, with logarithmic dependence on the number of LLM candidates and polynomial dependence on the nuisance estimate accuracy. Under a Tsybakov-type margin condition (Tsybakov, 2004), we derive faster regret rates that are minimax optimal. These results assume sufficiently small optimization error. We also prove a population "no-harm" property: optimal recalibration cannot reduce performance and yields a strict improvement when raw decisions are not perfectly calibrated. We give conditions under which estimation preserves this improvement. **(iv) Applications:** We apply ReCal-Route to an IMDb sentiment classification task and a real-world VLM safety screening task at a technology company. For both applications, ReCal-Route achieves the highest F_1 score in almost all settings.

2 Framework

2.1 Training data

We have a static dataset of n labeled items with binary label $Y \in \{0, 1\}$ and context vector $Z \in \mathcal{Z}$ observed before calling any LLM judge. In our VLM safety screening application, each item is an image; $Y = 1$ indicates unsafe content (e.g., violence), and Z contains dummy variables indicating which types of objects appear.

We assume access to a pool of J judges, each an LLM API. Calling judge j incurs cost $c_j \geq 0$ and returns a binary label $y_j \in \{0, 1\}$, a numeric confidence score $q_j \in [0, 1]$, and potentially other information. We combine all judge outputs into a vector x_j, whose domain is denoted by $\mathcal{X}$, and refer to it as a *prediction*. Absent other information, $x_j = (y_j, q_j)$. For each item in the dataset, we call each of the J judges to get their predictions.

Following Chen et al. (2020; 2024), we consider more than individual LLM APIs. We also include cascaded judges that combine predictions from multiple LLMs. For example, with two LLMs A and B, we can construct two additional judges, $A \wedge B$ and $A \vee B$, where $\wedge$ denotes AND and $\vee$ denotes OR. Their final answers are $\min\{y_A, y_B\}$ and $\max\{y_A, y_B\}$, respectively. Similarly, we can construct more complicated cascaded judges such as $(A \wedge B) \vee C$. Cascading has proved effective across a range of LLM applications (e.g., Dekoninck et al., 2025). For our routing problem, these simple Boolean operations substantially enrich the plan space without additional data collection costs: the labels of a cascaded judge can be computed directly from the labels already collected from its constituent judges.

To avoid confusion, we refer to any singleton or cascaded judge as a *plan*. Let $\mathcal{A}$ be a fixed finite collection of such call plans and c_a be the cost of calling plan $a \in \mathcal{A}$. Importantly, c_a can be much smaller than the total cost of all models involved in the plan. For example, for $a = A \wedge B$, $c_a = c_A$ when judge A outputs a negative label, since $\min\{y_A, y_B\} = 0$ regardless of judge B's output. Similarly, we denote by y_a and x_a the labels and whole prediction produced by plan a. When a involves multiple LLM judges, the dimensions of y_a and x_a can vary across items because of early

stopping. For example, when $a = A \wedge B$, y_a contains only judge A's label if $y_A = 0$, and both judges' labels otherwise.

2.2 Raw and Recalibrated Routing

The routing problem is to learn a *router* $\pi : \mathcal{Z} \to \mathcal{A}$ that selects a plan using the observed context to optimize the objective in Section 2.3, subject to $\mathbb{E}[c_{\pi(Z)}] \leq B$, where B is the available budget per item. Throughout the paper, we assume no distribution shift between training and deployment.

At deployment, for an item with context z, the router chooses plan $\pi(z)$, obtains the judge label $y_{\pi(z)}$, and incurs cost $c_{\pi(z)}$. We refer to this formulation as *raw routing*, as the prediction is dictated by the LLM judges. Using the judges' predictions as the final output is common practice in the LLM routing literature (e.g., Chen et al., 2020; 2024; Ong et al., 2025; Dekoninck et al., 2025).

While raw routing is standard for general LLM queries with unstructured responses, it need not be ideal for LLM judge routing, where the final output is a binary label. Using the training data described in Section 2.1, we can recalibrate the label y_a by fitting a prediction model for the ground truth Y using the context Z and the prediction X_a returned by each plan $a \in \mathcal{A}$. Specifically, we can fit a simple model, such as a linear model or a shallow neural network, to obtain an estimate $\widehat{p}_a(z, x)$ of $p_a^\star(z, x) \triangleq \mathbb{P}(Y = 1 \mid Z = z, X_a = x)$, and choose a threshold $\widehat{t}$ to convert the probability estimate into a binary predicted label.

For recalibrated routing, a *routing policy* $g = (\pi, p, t)$ combines a router π, a collection of conditional probability estimates $p = \{p_a : a \in \mathcal{A}\}$, where $p_a : \mathcal{Z} \times \mathcal{X} \to [0, 1]$, and a threshold $t \in [0, 1]$. We refer to (p, t) as nuisance parameters and write

$$h_{a,t}(z, x) = \mathbf{1}\{p_a(z, x) > t\} \tag{1}$$

for the final predicted label if plan a were chosen given context z and prediction x. We will justify the optimality of the thresholding rule in Step 3 of Section 3.

At deployment, for an item with context z and prediction x returned by the selected plan $\pi(z)$, the routing policy $g = (\pi, p, t)$ replaces the raw label $y_{\pi(z)}$ with

$$\widehat{Y}_g = h_{\pi(Z),t}(Z, X_{\pi(Z)}). \tag{2}$$

The rule in (2) can leave reliable raw labels largely unchanged. When the raw labels are poorly calibrated for some contexts or types of predictions, such as positive labels with low confidence, recalibration can substantially improve the final decisions without collecting more data or making additional LLM calls.

2.3 Classification metrics

With the definition of $\widehat{Y}_g$ in (2), we can define the classification metrics as follows:

$$\begin{aligned} \mathrm{TP}(g) &= \mathbb{E}[Y\widehat{Y}_g], & \mathrm{FP}(g) &= \mathbb{E}[(1-Y)\widehat{Y}_g], \\ \mathrm{FN}(g) &= \mathbb{E}[Y(1-\widehat{Y}_g)], & \mathrm{TN}(g) &= \mathbb{E}[(1-Y)(1-\widehat{Y}_g)], \end{aligned}$$

where $\mathrm{TP}, \mathrm{FP}, \mathrm{FN}, \mathrm{TN}$ stand for true positive, false positive, false negative and true negative. Following the standard practice, we focus on the F_β objective, for $\beta > 0$:

$$F_\beta(g) = \frac{(1+\beta^2)\mathrm{TP}(g)}{\mathbb{P}(\widehat{Y}_g = 1) + \beta^2 \mathbb{P}(Y = 1)}. \tag{3}$$

The F_β metric unifies precision as $\beta \to 0$, recall as $\beta \to \infty$, and the F_1 score at $\beta = 1$. Given nuisance parameters (p, t), budget-constrained recalibrated routing can be formulated as a population optimization problem:

$$\max_{\pi} \; F_\beta(\pi, p, t) \quad \text{s.t.} \quad \mathbb{E}[c_{\pi(Z)}] \leq B. \tag{4}$$

In contrast to the objectives considered in existing work (e.g., Chen et al., 2024; Hu et al., 2024; Ong et al., 2025; Dekoninck et al., 2025; Zhang et al., 2025; Li et al., 2026), the F_β objective is a nonlinear functional of the distribution of (Y, Z, X): it is a ratio of moments, including $\mathrm{TP}(g)$ and $\mathbb{P}(\widehat{Y}_g = 1)$, as shown in (3).

3 METHOD

In this section, we describe ReCal-Route, which consists of six steps: (1) forming a partition of $\mathcal{Z}$ such that the router depends on Z only through the cell it belongs to; (2) estimating the nuisance parameters $p = \{p_a : a \in \mathcal{A}\}$; (3) solving for the optimal threshold at a given value of auxiliary variable ϕ; (4) optimizing the router via an MILP; (5) updating ϕ and selecting the best feasible iterate; and (6) refitting the conditional probabilities and threshold using full data.

Step 1: context partition. To control routing complexity and improve interpretability, we learn a coarse partition χ of context Z into K cells, assigning one plan per cell: $\pi(z) = \pi_{\chi(z)}$. This gives a finite binary parametrization of the router (Step 4). We use quantile bins of a learned scalar summary of Z, as detailed in Appendix D.2.

Step 2: cross-fitting conditional probabilities. For item i in the training set, we have access to $(Y_i, Z_i, (X_{ia}, c_{ia})_{a\in\mathcal{A}})$. To reduce bias in estimating $p_a^\star(z, x)$, we split the data into V folds. For each held-out fold v, we use only items outside v to fit prediction models for Y using Z and each plan's prediction X_a. We fix a single loss ℓ, prediction model, and regularization choice. For example, we use linear regression with squared loss or logistic regression with logistic loss, both with ridge regularization. Write $\widehat{p}_a^{(-v)}$ for the resulting estimator for plan a. For each $i \in v$, compute $\widehat{p}_{ia}^{(-v)} = \widehat{p}_a^{(-v)}(Z_i, X_{ia})$.

Step 3: solving for the optimal threshold. Optimizing (3) is a fractional optimization problem where a common solution is linearization. Set $\rho = \mathbb{P}(Y = 1) > 0$. For a trial level ϕ, and fixed plan-specific binary decision rules $h = \{h_{a,t} : a \in \mathcal{A}\}$ in the form of (1), the transformed objective is

$$Q_\phi(\pi) = \mathbb{E}\big[\big((1+\beta^2)Y - \phi\big)h_{\pi(Z),t}(Z, X_{\pi(Z)})\big] - \phi\beta^2\rho. \tag{5}$$

For fixed h, (3) and Dinkelbach (1967) give

$$\max_{\pi:\,\mathbb{E}[C_{\pi(Z)}]\le B} Q_{\phi^\star}(\pi) = 0. \tag{6}$$

Here, $\phi^\star$ is the maximal attainable F_β score among routers satisfying the budget constraint, with h held fixed. For any fixed ϕ and router π, conditional on $(Z, X_a) = (z, x)$ with $a = \pi(z)$, predicting 1 rather than 0 changes the conditional contribution to (5) by $(1+\beta^2)p_a^\star(z, x) - \phi$. Thus, with $t = \phi/(1+\beta^2)$, the Bayes decision rule is

$$h_{a,t}^\star(z, x) = \mathbf{1}\left\{p_a^\star(z, x) > \frac{\phi}{1+\beta^2}\right\}. \tag{7}$$

When $\widehat{p} = p^\star$, (7) gives the optimal threshold $t = \phi/(1+\beta^2)$ for the transformed objective. At $\phi = \phi^\star$, (6) connects this rule to the optimal F_β value. Therefore, we set $\widehat{t} = \phi/(1+\beta^2)$ for any given ϕ. We will discuss how to choose ϕ in Step 5.

Step 4: optimizing router and threshold via MILPs. Fix the cross-fitted probabilities, a partition χ, and a trial metric value ϕ. Write $k_i = \chi(Z_i)$, set $\widehat{t} = \phi/(1+\beta^2)$ as in Step 3. We can estimate the F_β score and expected cost for a routing policy $g = (\pi, \widehat{p}, \widehat{t})$ with $d_{ia} = \mathbf{1}\{\widehat{p}_{ia}^{(-v)} > \widehat{t}\}$:

$$\widehat{F}(g) = \frac{(1+\beta^2)\sum_i Y_i d_{i,\pi(Z_i)}}{\sum_i d_{i,\pi(Z_i)} + \beta^2\sum_i Y_i}, \qquad \widehat{C}(g) = \frac{1}{n}\sum_i c_{i,\pi(Z_i)}. \tag{8}$$

Binary variables x_{ka} assign one plan to each occupied group. The assignment problem is

$$\begin{aligned} \widehat{Q}^\star(\phi) = \max_x \quad & \frac{1}{n}\sum_{i,a} d_{ia}\{(1+\beta^2)Y_i - \phi\}x_{k_i a} - \phi\beta^2\widehat{\rho} \\ \text{s.t.} \quad & \frac{1}{n}\sum_{i,a} c_{ia}x_{k_i a} \le B, \quad \sum_{a\in\mathcal{A}} x_{ka} = 1, \quad k = 1, \dots, K, \quad x_{ka} \in \{0, 1\}, \end{aligned} \tag{9}$$

where $\widehat{\rho} = n^{-1}\sum_i Y_i > 0$. For fixed ϕ, (9) is a mixed-integer linear program. Its solution gives a router that maps each occupied cell k to the plan a for which $x_{ka} = 1$.

Step 5: choosing the trial level. If we could solve (9) for every trial value ϕ, we could seek an empirical solution to (3) by finding a value satisfying the sample analogue of (6). We analyze the

theoretical properties of this version in Section 4. In practice, to avoid a costly grid search, we use an iterative algorithm in the implementation. Starting from $\phi = 0$, update ϕ to $\widehat{F}(g)$, recompute decisions, and repeat until a stopping criterion is met.

Step 6: refitting conditional probabilities and threshold. Let $\widehat{\pi}$ be the router retained in Step 5. We refit the same prediction model, using loss ℓ and the same regularization choice, on all training data to obtain $\widehat{p} = \{\widehat{p}_a : a \in \mathcal{A}\}$. With $\widehat{\pi}$ and $\widehat{p}$ fixed, we choose the final threshold $\widehat{t}$ to maximize training F_β using the refitted probabilities $\widehat{p}_{\widehat{\pi}(Z_i)}(Z_i, X_{i,\widehat{\pi}(Z_i)})$ directly. The final policy is $\widehat{g} = (\widehat{\pi}, \widehat{p}, \widehat{t})$.

Algorithm 1 summarizes ReCal-Route. We can consider multiple partitions in Step 1 and loss functions/regularization levels in Step 2, selecting the policy with the highest estimated F_β. The upper bounds in Section 4 extend to a fixed finite set of options if their assumptions hold uniformly and the partitions are fixed independently of the analyzed sample; partitions learned on that sample require additional control. For clarity, we present the algorithm with one partition, loss, and regularization choice. Our experiments (Section 5) use multiple partitions from Appendix D.2 and linear and logistic regressions with ridge regularization for recalibration.

Algorithm 1 ReCal-Route: Budget Constrained Recalibrated Routing of LLM judges

Input: training data $\{(Y_i, Z_i, (X_{ia}, c_{ia})_{a\in\mathcal{A}})\}_{i=1}^n$; a loss ℓ, prediction model, and regularization choice; K cells; V folds; $\beta > 0$; budget B; stopping rule.

Training phase

Learn a partition $\chi : \mathcal{Z} \to \{1, \dots, K\}$ as in Step 1; set $k_i = \chi(Z_i)$.
Split the training indices into V folds as in Step 2.
for each held-out fold v **do**
  For every plan $a \in \mathcal{A}$, fit $\widehat{p}_a^{(-v)}$ using $\{(Y_j, Z_j, X_{ja}) : j \notin v\}$ with the fixed loss ℓ, model, and regularization.
  Compute $\widehat{p}_{ia}^{(-v)} = \widehat{p}_a^{(-v)}(Z_i, X_{ia})$ for $i \in v$ and $a \in \mathcal{A}$.
end for
Initialize $\phi \leftarrow 0$.
repeat
  Set $\widehat{t} \leftarrow \phi/(1+\beta^2)$ as in Step 3.
  Set $d_{ia} \leftarrow \mathbf{1}\{\widehat{p}_{ia}^{(-v)} > \widehat{t}\}$ for each $i \in v$ and $a \in \mathcal{A}$.
  Solve the MILP (9) as in Step 4; set $\pi(z) = a$ for the plan with $x_{\chi(z),a} = 1$.
  Evaluate $g = (\pi, \widehat{p}, \widehat{t})$ using (8); retain the feasible pair $(\pi, \widehat{t})$ with the largest $\widehat{F}(g)$ seen so far.
  Update $\phi \leftarrow \widehat{F}(g)$ as in Step 5.
until the stopping rule is met.
Let $\widehat{\pi}$ be the retained router.
Refit the same learner on all training data to obtain $\widehat{p} = \{\widehat{p}_a : a \in \mathcal{A}\}$.
Set $a_i = \widehat{\pi}(Z_i)$ and $s_i = \widehat{p}_{a_i}(Z_i, X_{ia_i})$, $i = 1, \dots, n$.
$\widehat{t} \in \arg\max_{t\in\mathbb{R}} \dfrac{(1+\beta^2)\sum_i Y_i \mathbf{1}\{s_i > t\}}{\sum_i \mathbf{1}\{s_i > t\} + \beta^2 \sum_i Y_i}$.
return $(\chi, \widehat{\pi}, \widehat{p}, \widehat{t})$.

Deployment phase

For context z, execute plan $a = \widehat{\pi}(z)$ to obtain prediction x_a, and return $\mathbf{1}\{\widehat{p}_a(z, x_a) > \widehat{t}\}$.

4 THEORY

We analyze ReCal-Route (Algorithm 1) with a loss function ℓ, a fixed regularization choice, and a fixed partition $\chi : \mathcal{Z} \to \{1, \dots, K\}$ chosen independently of the training observations. Let Π be the routers that assign one plan to each cell, so $|\Pi| \le |\mathcal{A}|^K$.

Assumption 1. *The observations satisfy* $\big(Y_i, Z_i, (X_{ia}, c_{ia})_{a\in\mathcal{A}}\big)_{i=1}^n \overset{i.i.d.}{\sim} \big(Y, Z, (X_a, C_a)_{a\in\mathcal{A}}\big)$*, with* $\rho = \mathbb{P}(Y = 1) \ge \rho_0 > 0$ *and costs bounded by* $c_{\max}$.

For each plan $a \in \mathcal{A}$, we model $p_a^\star(z, x)$ using its context z and returned prediction x, as in Step 2. Let $\Phi_a(z, x)$ be a fixed, finite-dimensional numerical feature map, including intercept terms, and let

θ be its coefficient vector. With a nondecreasing link $\psi : \mathbb{R} \to [0, 1]$, the model is

$$p_a(z, x; \theta) = \psi\{\Phi_a(z, x)^\top \theta\}. \tag{10}$$

Equation (10) includes clipped linear and logistic probability models. For a cascade, Φ_a can use separate blocks for the possible collections of returned judge predictions, setting unused blocks to zero. This represents the singleton and pair models in Section 5 without requiring X_a to have a fixed length. Coefficients may be shared across plans that use the same fitted model.

4.1 Regret bound: slow rate under weaker assumptions

With the fitted probabilities fixed, we compare routers $\pi \in \Pi$ and cutoffs $t \in \mathbb{R}$ using $\widehat{F}(g)$ and $\widehat{C}(g)$ from (8). Write $C(g) = \mathbb{E}[c_{\pi(Z)}]$ for the population cost. Let $\widehat{\theta}_a^{(-v)}$ denote the coefficients of the estimates $\widehat{p}_a^{(-v)}$ from Step 2, with $v = 0$ representing the full-data estimate $\widehat{p}_a$ from Step 6.

Assume there exists some pseudo-parameter $\theta_a^\star$ such that $\widehat{p}_a^{(-v)}$ converges to

$$\bar{p}_a(z, x) = p_a(z, x; \theta_a^\star). \tag{11}$$

The population model in (11) may differ from $p_a^\star$. Write

$$\bar{p}_\pi = \bar{p}_{\pi(Z)}(Z, X_{\pi(Z)}), \qquad \widehat{p}_\pi^{(-v)} = \widehat{p}_{\pi(Z)}^{(-v)}(Z, X_{\pi(Z)}). \tag{12}$$

Define the error of the conditional probability estimates as follows.

$$a_n = \sup_{\pi \in \Pi} \max_{0 \le v \le V} \|\widehat{p}_\pi^{(-v)} - \bar{p}_\pi\|_\infty. \tag{13}$$

We also define the oracle analogue of $\bar{p}_\pi$:

$$p_\pi := p^\star_{\pi(Z)}(Z, X_{\pi(Z)}) = \mathbb{P}(Y = 1 \mid Z, X_{\pi(Z)}). \tag{14}$$

Let $V_\pi^\star = \sup_d F_\beta\{d(Z, X_{\pi(Z)})\}$ over all binary predictions d. Put $\tau_\pi = V_\pi^\star/(1 + \beta^2)$.

Assumption 2. *For every router $\pi \in \Pi$, there is a cutoff $\bar{\tau}_\pi$ such that*

$$\mathbf{1}\{\bar{p}_\pi > \bar{\tau}_\pi\} = \mathbf{1}\{p_\pi > \tau_\pi\} \quad \textit{almost surely}. \tag{15}$$

Assumption 2 permits misspecified probabilities if thresholding $\bar{p}_\pi$ recovers the oracle decisions.

The next assumption controls the margins $S_\pi = |p_\pi - \tau_\pi|$ and $\bar{S}_\pi = |\bar{p}_\pi - \bar{\tau}_\pi|$. This is analogous to the margin condition in the classification literature, which requires the mass around the Bayes classifier's decision boundary to be small (Tsybakov, 2004; Yan et al., 2018; Chzhen, 2020).

Assumption 3. *For any $\pi \in \Pi$, fixed constants $L, r_0, \alpha, \nu > 0$ and $p > 1$ satisfy, for all $0 < r \le r_0$ and $v \ge 1$,*

$$\mathbb{P}(S_\pi \le r) \le L r^\alpha, \quad \mathbb{P}(\bar{S}_\pi \le r) \le L r^{\alpha/\nu}, \quad \mathbb{P}(\bar{S}_\pi \le r,\ S_\pi > v r^{1/\nu}) \le L r^{\alpha/\nu} v^{-p}. \tag{16}$$

The third inequality in Eq. (16) may be less intuitive. It is satisfied, for example, when the margins follow the power law $\bar{S}_\pi \asymp S_\pi^\nu$, together with the first inequality. This gives a simple interpretation of ν: under this relation, when $\nu = 1$, the population score $\bar{p}_\pi$ and the true probability p_π lie at comparable distances from their respective cutoffs. When $\nu > 1$, $\bar{p}_\pi$ can be much more compressed toward its cutoff $\bar{\tau}_\pi$ than p_π is toward τ_π. Thus, vanishing F_β regret remains possible with imperfect recalibration, provided that decisions align and the margins shrink together.

Let $D = \max_a \dim \Phi_a$. For $\delta \in (0, 1)$ and sufficiently large C_0, set

$$e_n(\delta) = C_0 \sqrt{\frac{KD + K \log|\mathcal{A}| + \log(8V/\delta)}{n}},\ b_n(\delta) = c_{\max} \sqrt{\frac{K \log|\mathcal{A}| + \log(8/\delta)}{2n}}. \tag{17}$$

The following theorem establishes a regret bound. The proof is in Appendix A.2.

Theorem 1. *Suppose Assumptions 1, 2, and 3 hold. Run Algorithm 1 with budget $B - b_n(\delta)$ and let $\widetilde{g}$ be the output. For sufficiently large n, with probability at least $1 - \delta$,*

$$C(\widetilde{g}) \le B, \qquad F_\beta(\widetilde{g}) \ge \max_{g=(\pi,p,t): C(g) \le B - 2b_n(\delta)} V_\pi^\star - C\{e_n(\delta) + a_n^{(1+\alpha)/\nu}\}, \tag{18}$$

for a constant $C < \infty$ independent of n and δ, provided the budget constraint is feasible.

4.2 REGRET BOUND: FAST RATE UNDER STRONGER ASSUMPTIONS

For each $\pi \in \Pi$, let C_π denote the expected cost. Recall the definition of $V_\pi^\star$ in Section 4.1. Set

$$\Delta_\pi = \max_{\pi' \in \Pi:\, C_{\pi'} \le B} V_{\pi'}^\star - V_\pi^\star.$$

Let $V_n = \max_{\pi \in \Pi:\, C_\pi \le B} V_\pi^\star$ denote the best population value among budget-feasible routers. The subscript n allows the underlying distribution, and hence these values, to vary with sample size. The next assumption restricts the gaps of suboptimal routing policies. It states that any suboptimal policy has either a constant or an $O(n^{-\gamma})$ gap from the global optimum. The assumption is in similar spirit to margin conditions in the econometrics literature (e.g., Ji et al., 2026; Semenova, 2025).

Assumption 4. *Let $\Pi_0 \subset \Pi$ be fixed and nonempty. For some fixed $\kappa > 0$, $C_\pi \le B - \kappa$ for every $\pi \in \Pi_0$. There exist constants $\gamma, c, c_\Delta, C_\Delta > 0$ such that every $\pi \in \Pi_0$ satisfies,*

$$\Delta_\pi = 0 \quad \textit{or} \quad c_\Delta n^{-\gamma} \le \Delta_\pi \le C_\Delta n^{-\gamma}. \tag{19}$$

Every $\pi \in \Pi \setminus \Pi_0$ with $C_\pi \le B$ satisfies $\Delta_\pi \ge c$.

The uniform value-error bound in Theorem 1 can overstate routing regret: well-separated routers remain correctly ranked, while selecting a nearly tied router incurs little loss. The next theorem establishes a fast rate under Assumption 4. The proof is given in Appendix A.4.

Theorem 2. *Suppose Assumptions 1, 2, 3, and 4 hold and $a_n = o_p(1)$. Run Algorithm 1 with budget $B - b_n(\delta_n)$, where $\delta_n \downarrow 0$ and $\log(1/\delta_n) = o(n)$. With probability tending to one, the algorithm returns a policy $\widetilde{g}$ with selected router $\widetilde{\pi} \in \Pi_0$ with $C(\widetilde{g}) \le B$, and*

$$V_n - F_\beta(\widetilde{g}) = \begin{cases} O_p\Big(n^{-(1+\alpha)/(2+\alpha)} + a_n^{(1+\alpha)/\nu}\Big), & 0 < \gamma < 1/2, \\ O_p\Big(n^{-(1+\alpha)/(2+\alpha)} + a_n^{(1+\alpha)/\nu} + n^{-\gamma}\Big), & \gamma \ge 1/2. \end{cases} \tag{20}$$

In Lemma 1 in Appendix A.1, we provide conditions under which $a_n = O_p(n^{-1/2})$. In this case, (20) reduces to $O_p(n^{-r_{\alpha,\nu,\gamma}})$, where

$$r_{\alpha,\nu,\gamma} = \begin{cases} \min\left\{\dfrac{1+\alpha}{2+\alpha}, \dfrac{1+\alpha}{2\nu}\right\}, & 0 < \gamma < 1/2, \\ \min\left\{\gamma, \dfrac{1+\alpha}{2+\alpha}, \dfrac{1+\alpha}{2\nu}\right\}, & \gamma \ge 1/2. \end{cases} \tag{21}$$

The next theorem gives a corresponding minimax lower bound. The proof is given at the end of Appendix C.

Theorem 3. *Fix $\alpha, \nu, \gamma > 0$ and $p > 1$. There exist a fixed partition χ, plan collection $\mathcal{A}$, and recalibration model with loss ℓ for which the following holds. Fit recalibration scores $\widehat{p}$ via empirical minimization of ℓ on the full training sample. Let $\mathcal{P}$ be the class of all distributions satisfying Assumptions 1–3. For some $c, c_0 > 0$ and all sufficiently large n,*

$$\inf_{(\widehat{\pi},\widehat{t})} \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\big\{V_n - F_{\beta,\mathbb{P}}(\widehat{\pi}, \widehat{p}, \widehat{t}) \ge c n^{-r_{\alpha,\nu,\gamma}}\big\} \ge c_0. \tag{22}$$

The infimum is over all data-dependent router and cutoff selections $(\widehat{\pi}, \widehat{t}) \in \Pi \times \mathbb{R}$, and $r_{\alpha,\nu,\gamma}$ is given in (21).

4.3 POPULATION GAIN OVER RAW ROUTING

Let r_a denote plan a's raw binary decision rule. Fix a router π and let $d_\pi^{\mathrm{raw}} = r_{\pi(Z)}(X_{\pi(Z)})$ be its raw decision, and $\tau_\pi^\star = V_\pi^\star/(1+\beta^2)$. We prove at the start of Appendix A.3 that

$$V_\pi^\star - F_\beta(d_\pi^{\mathrm{raw}}) = \frac{(1+\beta^2)\,\mathbb{E}[|p_\pi - \tau_\pi^\star| \mathbf{1}\{d_\pi^{\mathrm{raw}} \ne \mathbf{1}\{p_\pi > \tau_\pi^\star\}\}]}{\mathbb{E}[d_\pi^{\mathrm{raw}}] + \beta^2 \rho}. \tag{23}$$

Thus, recalibration improves performance where the raw decision lies on the wrong side of the optimal threshold and disagreements farther away increase the gain.

Define the population gain over raw routing by

$$\Delta^{\mathrm{raw}} = \max_{\pi \in \Pi:\, C_\pi \le B} V_\pi^\star - \max_{\pi \in \Pi:\, C_\pi \le B} F_\beta(d_\pi^{\mathrm{raw}}) \ge 0. \tag{24}$$

The inequality in (24) holds because post-call decision rules include the raw decisions. Let $\widehat{g}_{\mathrm{raw}}$ be the policy learned by raw routing using the same training data, router class Π, and budget $B - b_n(\delta_n)$. Under our conditions, the recalibrated router improves over the raw router with high probability, evaluated on population F_β. The proof is given in Appendix A.3.

Theorem 4. *Under the conditions of Theorem 2, if* $\Delta_0 = \liminf_n \Delta^{\mathrm{raw}} > 0$*, then*

$$\mathbb{P}\{F_\beta(\widetilde{g}) - F_\beta(\widehat{g}_{\mathrm{raw}}) \ge \Delta_0/2\} \to 1.$$

5 Applications

We apply ReCal-Route (Algorithm 1) to IMDb sentiment classification and VLM safety screening at a tech company, optimizing F_1 in both applications. To account for splitting uncertainty, we report mean test F_1 over multiple data splits. Appendices D and E give experimental details.

Judges and plans. We consider four judges for the IMDb application and six judges for the VLM application. The collection of plans includes all singleton judges and all two-judge cascades formed using $\wedge$ or $\vee$, with both calling orders included. Thus, $|\mathcal{A}| = 28$ for IMDb and $|\mathcal{A}| = 66$ for VLM. To limit the number of conditional probability models, we fit only $\widehat{p}_j$ for each judge $j = 1, \ldots, J$ and $\widehat{p}_{jk}$ for each unordered pair $1 \le j < k \le J$. Each model uses the context and the outputs of the corresponding judge or pair of judges. For a cascade that calls A before B, we use $\widehat{p}_A$ for $\widehat{p}_{A \wedge B}$ when $y_A = 0$ and for $\widehat{p}_{A \vee B}$ when $y_A = 1$. In both cases, the plan stops after calling A; otherwise, we use the model fitted for the pair $\{A, B\}$.

Competing methods. We also include the raw router, which omits recalibration. We compare with F_1 adaptations of RouterBench-kNN (Hu et al., 2024), Avengers-Pro (Zhang et al., 2025), Unified Router (Dekoninck et al., 2025), WISERouter (Li et al., 2026), and a depth-two cascade (Dekoninck et al., 2025). The IMDb application also includes FrugalML (Chen et al., 2020).

5.1 IMDb sentiment classification

The task is to predict whether a movie review expresses positive ($Y = 1$) or negative ($Y = 0$) sentiment. We use 25,000 IMDb reviews from the public FrugalML and four judges: Google, Baidu, Amazon, and VADER. We consider three types of contexts: 8 manually generated features, 64-dimensional TF–IDF/LSA features, or their concatenation. The budget B is measured in nominal cost units (e.g., dollars) per 10,000 reviews. We use 40 grouped 80/20 train/test splits. Figure 1 shows F_1 against the expected budget. Averaging over each budget value and data splits, ReCal-Route improves upon competing methods by 3.17 percentage points (pp) with LSA context and 3.22 pp with combined context.

5.2 VLM safety screening

The task is to identify AI-generated 3D content that violates the moderation policy, using human moderation labels as the ground truth ($Y = 1$ for a violation and $Y = 0$ otherwise). We study 3,788 items and 6 VLM judges under three input settings: separate rendered views (render-only), a single composite image of those views (render–mesh grid), and rendered views together with the creator's text prompt (multimodal). The dataset substantially oversamples unsafe content and was curated for model improvement. Before querying judges, we use 6 content tags curated by our collaborating tech company as context. We also generate context using an off-the-shelf CLIP embedding (Radford et al., 2021; Cherti et al., 2023). The budget B is measured in nominal cost units (e.g., dollars) per 1,000 items. We evaluate each setting on 100 stratified 70/30 development/test splits. Tables 1 and 2 compare methods at $B = 2.0$ for two types of context. ReCal-Route outperforms all competing methods in five of the six settings and ranks second for the multimodal task with CLIP embeddings.

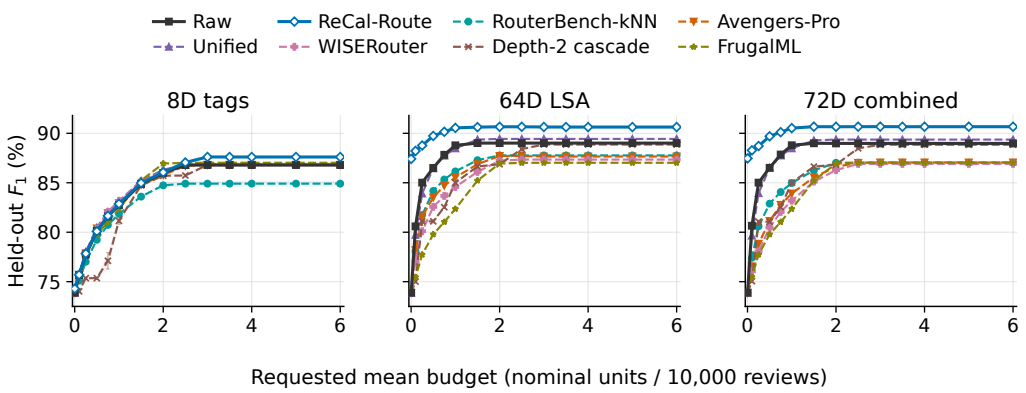


Figure 1: IMDb F_1 versus expected budget across three contexts (40 grouped splits). Bands are descriptive pointwise 95% Student-t intervals. Full curves, including realized-expenditure comparisons, are in Appendix E.2.

Table 1: VLM mean held-out F_1 (%) with 6 content tags at expected budget $B = 2.0$ nominal cost units (e.g., dollars) per 1,000 items. All entries use 100 splits. Bold marks the highest mean in each column.

Method	Render-only	Render–mesh grid	Multimodal
Raw	81.77	84.33	75.72
ReCal-Route	**85.19**	**85.63**	**77.15**
RouterBench-kNN	75.13	77.66	71.64
Avengers-Pro	78.73	81.01	74.31
Unified Router	80.58	83.34	76.37
WISERouter	81.37	84.87	75.82
Depth-2 cascade	81.02	81.40	75.55

Table 2: VLM mean held-out F_1 (%) with CLIP embeddings. Other settings are the same as Table 1. Dashes indicate that Unified Router and the depth-two cascade fail to produce any feasible router in 14% and 20% of the multimodal splits, respectively.

Method	Render-only	Render–mesh grid	Multimodal
Raw	81.38	82.93	73.09
ReCal-Route	**85.47**	**84.99**	76.95
RouterBench-kNN	78.41	80.59	**77.86**
Avengers-Pro	82.07	82.78	75.54
Unified Router	75.19	76.40	-
WISERouter	81.78	82.38	73.58
Depth-2 cascade	78.44	80.22	-

6 Conclusion and Limitations

We formulate routing for LLM judges as constrained nonlinear policy learning that directly optimizes F_β under a budget. ReCal-Route combines cross-fitted conditional probability estimates with MILP-based routing and threshold selection. Recalibration uses context and returned judge outputs without additional calls. Under the stated assumptions, we establish slow and fast regret rates, minimax optimality, and a no-harm property of population recalibration. IMDb sentiment classification and VLM safety screening show F_1 gains over raw routing and competing methods across budgets and contexts. We assume outputs from every candidate judge for every training item and no distribution shift between training and deployment. Extending the method to partially observed training data and deployment-time distribution shift remains for future work.

AI USE STATEMENT

We use generative AI tools to assist with theory and proof drafts, prose and notation editing, LaTeX, and benchmark code. The authors reviewed all AI-assisted work and take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

For theoretical results, statements are stated in the main text and proofs are provided in the Appendix. The code for reproducing the experiments is contained in a .zip file along with this submission.

ETHICS STATEMENT

The application concerns sensitive-content moderation. Any release must protect creator and annotator privacy, document how moderator votes and label corrections were produced, and state the permitted use of proprietary items and model outputs. Automated routing supports rather than replaces the governance and appeal procedures required for high-impact moderation decisions.

REFERENCES

S. Athey and S. Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

L. Chen, M. Zaharia, and J. Zou. FrugalML: How to use ML prediction APIs more accurately and cheaply. In *Advances in Neural Information Processing Systems*, volume 33, pages 10685–10696, 2020.

L. Chen, M. Zaharia, and J. Zou. FrugalGPT: How to use large language models while reducing cost and improving performance. *Transactions on Machine Learning Research*, 2024.

M. Cherti, R. Beaumont, R. Wightman, M. Wortsman, G. Ilharco, C. Gordon, C. Schuhmann, L. Schmidt, and J. Jitsev. Reproducible scaling laws for contrastive language–image learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

E. Chzhen. Optimal rates for nonparametric F-score binary classification via post-processing. *Mathematical Methods of Statistics*, 29:87–105, 2020.

P. L. Cohen and C. B. Fogarty. No-harm calibration for generalized Oaxaca–Blinder estimators. *Biometrika*, 111(1):331–338, 2024.

J. Dekoninck, M. Baader, and M. Vechev. A unified approach to routing and cascading for LLMs. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

W. Dinkelbach. On nonlinear fractional programming. *Management Science*, 13(7):492–498, 1967.

Y. Geifman and R. El-Yaniv. SelectiveNet: A deep neural network with an integrated reject option. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, 2017.

Q. J. Hu, J. Bieker, X. Li, N. Jiang, B. Keigwin, G. Ranganath, K. Keutzer, and S. K. Upadhyay. RouterBench: A benchmark for multi-LLM routing system. arXiv:2403.12031, 2024.

W. Ji, L. Lei, and A. Spector. Model-agnostic covariate-assisted inference on partially identified causal effects. arXiv:2310.08115v3, 2026.

W. Kotłowski and K. Dembczyński. Surrogate regret bounds for generalized classification performance metrics. In *Asian Conference on Machine Learning*, 2016.

O. O. Koyejo, N. Natarajan, P. K. Ravikumar, and I. S. Dhillon. Consistent binary classification with generalized performance metrics. In *Advances in Neural Information Processing Systems*, 2014.

L. Lei and P. Ding. Regression adjustment in completely randomized experiments with a diverging number of covariates. arXiv:1806.07585, 2018.

Y. Li, Z. Gao, and L. V. S. Lakshmanan. WISERouter: LLM routing with workload budget constraint. arXiv:2607.23765, 2026.

D. Madras, T. Pitassi, and R. Zemel. Predict responsibly: Improving fairness and accuracy by learning to defer. In *Advances in Neural Information Processing Systems*, 2018.

H. Mozannar and D. Sontag. Consistent estimators for learning to defer to an expert. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.

H. Narasimhan, H. Ramaswamy, A. Saha, and S. Agarwal. Consistent multiclass algorithms for complex performance measures. In *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

H. Narasimhan, A. Cotter, and M. Gupta. Optimizing generalized rate metrics with three players. In *Advances in Neural Information Processing Systems*, 2019.

I. Ong, A. Almahairi, V. Wu, W.-L. Chiang, T. Wu, J. E. Gonzalez, M. W. Kadous, and I. Stoica. RouteLLM: Learning to route LLMs with preference data. In *The Thirteenth International Conference on Learning Representations*, 2025.

S. A. Puthiya Parambath, N. Usunier, and Y. Grandvalet. Optimizing F-measures by cost-sensitive classification. In *Advances in Neural Information Processing Systems*, 2014.

A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.

V. Semenova. Debiased machine learning of aggregated intersection bounds and other causal parameters. arXiv:2303.00982v3, 2025.

A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

A. van der Vaart and J. A. Wellner. A local maximal inequality under uniform entropy. *Electronic Journal of Statistics*, 5:192–203, 2011. `https://doi.org/10.1214/11-EJS605`.

B. Yan, S. Koyejo, K. Zhong, and P. Ravikumar. Binary classification with karmic, threshold-quasi-concave metrics. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 5531–5540, 2018.

Y. Zhang, H. Li, J. Chen, H. Zhang, P. Ye, L. Bai, and S. Hu. Beyond GPT-5: Making LLMs cheaper and better via performance-efficiency optimized routing. In *DAI '25: Proceedings of the 2025 7th International Conference on Distributed Artificial Intelligence*, pages 122–129. ACM, 2025.

A Supplementary Details and Proofs

A.1 Regularity conditions

For plan a, let $\ell_a(\theta)$ denote the per-item training loss induced by the fixed loss ℓ and its regularization. Its argument is the complete training record. If the plan uses several regression blocks, ℓ_a sums their losses, each evaluated using its judge predictions from that record; each block therefore uses every training item, including items on which the plan would stop before revealing it. The V folds are fixed independently of the observations, with V fixed and each training subset of size proportional to n.

We make the following assumption.

Assumption 5.

(i) ***Features and coefficients.*** *The features satisfy $\|\Phi_a(Z, X_a)\|_2 \le M_\Phi$ a.s., and coefficient domains Θ_a are fixed, compact, and convex.*

(ii) ***Lipschitz link.*** *The link ψ is Lipschitz.*

(iii) ***Differentiability.*** *Each per-item loss ℓ_a is differentiable in θ, with $\nabla Q_a = \mathbb{E}[\nabla_\theta \ell_a]$ for $Q_a(\theta) = \mathbb{E}[\ell_a(\theta)]$.*

(iv) ***Population curvature.*** *The population objectives Q_a are γ_0-strongly convex for a common $\gamma_0 > 0$.*

(v) ***Bounded Lipschitz gradients.*** *There are constants $G_\ell, L_\ell > 0$, independent of a and n, such that, almost surely, for all $a \in \mathcal{A}$ and $\theta, \theta' \in \Theta_a$,*

$$\begin{aligned} \|\nabla_\theta \ell_a(\theta)\|_2 &\le G_\ell, \\ \|\nabla_\theta \ell_a(\theta) - \nabla_\theta \ell_a(\theta')\|_2 &\le L_\ell \|\theta - \theta'\|_2. \end{aligned} \tag{25}$$

Under the sampling condition in Assumption 1 and Assumption 5, we can prove the following lemma.

Lemma 1. *Under the sampling condition in Assumption 1 and Assumption 5, suppose $\theta_a^\star$ minimizes Q_a and each $\widehat{\theta}_a^{(-v)}$ minimizes the corresponding empirical average of ℓ_a. Then $a_n = O_p(n^{-1/2})$.*

The proof is in Appendix A.2.

Notation and the theoretical search. For a router π, write $U_\pi = (Z, X_{\pi(Z)})$ for its returned features. Index the fixed router class as $\Pi = \{\pi_j : j \in \mathcal{J}\}$ and abbreviate $U_j = U_{\pi_j}$, $p_j = p_{\pi_j}$, $\bar{p}_j = \bar{p}_{\pi_j}$, $\bar{\tau}_j = \bar{\tau}_{\pi_j}$, and $\bar{S}_j = \bar{S}_{\pi_j}$. Likewise, $\widehat{p}_j^{(-v)} = \widehat{p}_{\pi_j}^{(-v)}$. The selected index $\widehat{j}$ denotes the router $\widehat{\pi} = \pi_{\widehat{j}}$, and $\widehat{g}$ is the final policy denoted by $\widehat{g}$ in Step 6. Because costs depend only on the router, write $C(\pi) = C_\pi = C(g)$ and $\widehat{C}(\pi) = \widehat{C}(g)$ for any policy $g = (\pi, p, t)$. To describe the shrinking gaps in Assumption 4, allow the population distribution to vary with n, writing it as P_n. Put $\mathcal{J}_0 = \{j : \pi_j \in \Pi_0\}$ and

$$V_{j,n} = V_{\pi_j}^\star, \qquad C_{j,n} = C_{\pi_j}, \qquad V_n = \max_{j : C_{j,n} \le B} V_{j,n}, \qquad \Delta_{j,n} = V_n - V_{j,n} = \Delta_{\pi_j}.$$

Thus V_n in (20) is the best population value attainable at budget B. Constants in the assumptions are uniform in n. Throughout Appendix A, $Pf = \mathbb{E}_{P_n}[f]$ denotes population expectation, and $\mathbb{P}_N f$ denotes the average over the training subset of size N under discussion. For a fitted f, Pf evaluates that realized function on a fresh observation. The norm in (13) is the essential supremum under the law of U_π, with the fitted function held fixed. For $v = 0$, superscript $(-v)$ denotes the full-data fit, also written $\widehat{p}^{(0)}$ or $\widehat{p}$.

For the upper-bound arguments, take the theoretical search to maximize the cross-fitted criterion over all feasible router–cutoff pairs, followed by the full-data cutoff maximization in Step 6. This makes explicit the optimization-accuracy requirement mentioned in the Introduction. The sharp-rate proof also permits an empirical objective error $o_p(n^{-1/2})$. The practical iteration in Step 5 is described in Appendix D; its stopping rule alone does not guarantee this optimization accuracy.

A.2 Proofs of Lemma 1 and Theorem 1

Proof of Lemma 1. Fix a plan a and a training subset of size $N \geq cn$. Write $\widehat{Q}_{N,a} = \mathbb{P}_N \ell_a$ and $Q_a = P\ell_a$. Suppress the fold index and let $\widehat{\theta}_a$ minimize $\widehat{Q}_{N,a}$, while $\theta_a^\star$ minimizes Q_a. The sampling condition in Assumption 1 and Assumption 5 imply

$$Z_{N,a} := \sup_{\theta \in \Theta_a} \|(\mathbb{P}_N - P)\nabla_\theta \ell_a(\theta)\|_2 = O_p(N^{-1/2}).$$

For each coordinate, the gradient is bounded by G_ℓ and is L_ℓ-Lipschitz in θ, by Assumption 5(v). A Euclidean grid on the fixed compact set Θ_a therefore covers the gradient functions uniformly over observations; centering increases the envelope and Lipschitz constant by at most a factor of two. Consequently the centered coordinate gradient classes have a common bounded envelope G and polynomial covering numbers: for constants A, v_0,

$$\sup_Q N(\epsilon \|G\|_{Q,2}, \mathcal{F}_a, L_2(Q)) \leq (A/\epsilon)^{v_0}, \qquad 0 < \epsilon \leq 1,$$

Here $N(\varepsilon, \mathcal{F}_a, L_2(Q))$ denotes a covering number, Q ranges over finitely supported probability measures, and $\mathcal{F}_a$ contains all centered coordinate gradients indexed by $\theta \in \Theta_a$. The entropy integral is finite. Symmetrization and the entropy bound give $\mathbb{E}^* Z_{N,a} \leq C/\sqrt{N}$, where $\mathbb{E}^*$ denotes outer expectation; bounded-difference concentration adds $C\sqrt{\log(1/\delta)/N}$ with failure probability at most δ.

Put $\Delta = \widehat{\theta}_a - \theta_a^\star$. Since Θ_a is convex, moving from either minimizer toward the other remains feasible. The directional derivative of the corresponding objective in this direction must be nonnegative, so

$$\langle \nabla Q_a(\theta_a^\star), \Delta \rangle \geq 0, \qquad \langle \nabla \widehat{Q}_{N,a}(\widehat{\theta}_a), \Delta \rangle \leq 0.$$

Strong convexity of Q_a gives the first inequality below; the two optimality conditions give the second, and Cauchy–Schwarz together with the definition of $Z_{N,a}$ gives the last:

$$\begin{aligned} \gamma_0 \|\Delta\|_2^2 &\leq \langle \nabla Q_a(\widehat{\theta}_a) - \nabla Q_a(\theta_a^\star), \Delta \rangle \\ &\leq \langle \nabla Q_a(\widehat{\theta}_a) - \nabla \widehat{Q}_{N,a}(\widehat{\theta}_a), \Delta \rangle \leq Z_{N,a} \|\Delta\|_2. \end{aligned}$$

Thus $\|\Delta\|_2 \leq Z_{N,a}/\gamma_0$. This argument uses population curvature; empirical strong convexity is unnecessary. Any deterministic regularization included in ℓ_a cancels from its empirical-minus-population gradient.

Bounded augmented features and the Lipschitz constant L_ψ of the link yield

$$\|\widehat{p}_a - \bar{p}_a\|_\infty \leq L_\psi M_\Phi \|\widehat{\theta}_a - \theta_a^\star\|_2.$$

A union bound over the fixed number of plans and fold/full-data fits therefore gives, with probability at least $1 - \delta$,

$$\max_{a,\, 0 \leq v \leq V} \|\widehat{p}_a^{(-v)} - \bar{p}_a\|_\infty \leq C\sqrt{\frac{1 + \log\{2|\mathcal{A}|(V+1)/\delta\}}{n}}. \tag{26}$$

The routed error a_n is bounded by this planwise error. Overlap between training subsets does not affect the union bound. □

Lemma 2 (Uniform control of fold and refitted values)**.** *Suppose Assumptions 1, 2, and 3 hold. Let I_v contain the indices in fold v. Put $d_{\pi,t}^{(-v)} = \mathbf{1}\{\widehat{p}_\pi^{(-v)} > t\}$ and $\omega_v = |I_v|/n$, and define the population value of the fold mixture by*

$$F^-_{\beta,\mathrm{mix}}(\pi, t) = \frac{(1+\beta^2)\sum_{v=1}^V \omega_v P(Y d_{\pi,t}^{(-v)})}{\sum_{v=1}^V \omega_v P d_{\pi,t}^{(-v)} + \beta^2 \rho}.$$

Let $\widehat{F}^0(\pi, t)$ and $F^0_\beta(\pi, t)$ be the empirical and population values of the full-data decision $d_{\pi,t}^{(0)}$. For sufficiently large C_0 in (17)*, with probability at least $1 - \delta$,*

$$\begin{aligned} \sup_{\pi,t} |\widehat{F}(\pi, \widehat{p}, t) - F^-_{\beta,\mathrm{mix}}(\pi, t)| &\leq e_n(\delta), \\ \sup_{\pi,t} |\widehat{F}^0(\pi, t) - F^0_\beta(\pi, t)| &\leq e_n(\delta), \\ \sup_\pi |\widehat{C}(\pi) - C(\pi)| &\leq b_n(\delta). \end{aligned} \tag{27}$$

Proof. For a fixed route, a threshold of a nondecreasing single-index score is an affine halfspace or a constant on each context cell, in the features $\Phi_{\pi(z)}(z, x)$ of the plan assigned to that cell. There are K disjoint cells, each using at most D coefficients. The deterministic class allowing all coefficient vectors and cutoffs therefore has VC dimension at most a constant times KD. It contains every fitted decision despite its data dependence. Apply bounded-VC concentration separately on each fixed fold and on the full sample, and take a union over the at most $|\mathcal{A}|^K$ routers that are constant on each cell of the fixed partition χ. The fold sizes are comparable and V is fixed, so, simultaneously for $h = d$ and $h = Yd$, the empirical means differ from their population counterparts by at most

$$C\sqrt{\{KD + K \log |\mathcal{A}| + \log(8V/\delta)\}/n}.$$

Uniformity permits substitution of every realized fold and full-data fit; independence of these fits is unnecessary. Weighting the fold bounds by ω_v controls the fold-mixture numerator and denominator. The same argument controls the full-data values. Concentration of prevalence ensures $\widehat{\rho} \geq \rho_0/2$, and the difference-of-ratios identity gives the first two inequalities of (27). Hoeffding's inequality over the finite route class gives the cost bound. Allocate fixed fractions of δ to these events and enlarge C_0. □

Proof of Theorem 1. For this proof, abbreviate $e_n = e_n(\delta)$ and $b_n = b_n(\delta)$. Write $w = 1 + \beta^2$, $\tau_\pi = V_\pi^\star/w$, $S_\pi = |p_\pi - \tau_\pi|$, and $d_\pi^\star = \mathbf{1}\{p_\pi > \tau_\pi\}$. Let $\bar{\tau}_\pi$ be the aligned population cutoff, and set $\bar{S}_\pi = |\bar{p}_\pi - \bar{\tau}_\pi|$. Integrating the two inequalities in (16) involving $\bar{S}_\pi$ gives

$$\begin{aligned}\mathbb{E}[S_\pi \mathbf{1}\{\bar{S}_\pi \leq r\}] &\leq r^{1/\nu}\mathbb{P}(\bar{S}_\pi \leq r) + r^{1/\nu}\int_1^\infty \mathbb{P}(\bar{S}_\pi \leq r, S_\pi > v r^{1/\nu})\, dv \\ &\leq L\left(1 + \frac{1}{p-1}\right) r^{(1+\alpha)/\nu}.\end{aligned} \tag{28}$$

This bound holds for $0 < r \leq r_0$ and extends to larger r with a larger constant, since $S_\pi \leq 1$. At the common cutoff $\bar{\tau}_\pi$, decision alignment and the definition of a_n imply, for every fold and the full-data refit,

$$\{d_{\pi,\bar{\tau}_\pi}^{(-v)} \neq d_\pi^\star\} \subseteq \{\bar{S}_\pi \leq a_n\}.$$

Consequently, all these decisions have weighted error at most $C a_n^{(1+\alpha)/\nu}$.

The Dinkelbach identity, proved in the next subsection, gives

$$V_\pi^\star - F_\beta(d) = \frac{wP[S_\pi \mathbf{1}\{d \neq d_\pi^\star\}]}{Pd + \beta^2\rho}.$$

Its fold-mixture version is

$$V_\pi^\star - F_{\beta,\mathrm{mix}}^-(\pi, t) = \frac{w\sum_v \omega_v P[S_\pi \mathbf{1}\{d_{\pi,t}^{(-v)} \neq d_\pi^\star\}]}{\sum_v \omega_v P d_{\pi,t}^{(-v)} + \beta^2\rho}.$$

Every mixture value is therefore at most $V_\pi^\star$. Using the aligned cutoff and the denominator lower bound $\beta^2\rho_0$ yields

$$\begin{aligned}\sup_t F_{\beta,\mathrm{mix}}^-(\pi, t) &\geq V_\pi^\star - C a_n^{(1+\alpha)/\nu}, \\ \sup_t F_\beta^0(\pi, t) &\geq V_\pi^\star - C a_n^{(1+\alpha)/\nu}.\end{aligned} \tag{29}$$

On the event in Lemma 2, every comparator route with $C(\pi) \leq B - 2b_n$ is empirically feasible and every returned route satisfies $C(\widehat{\pi}) \leq B$. Empirical selection, the first bound in (29), and the upper bound of each mixture by its route oracle imply

$$\max_{\pi : C(\pi) \leq B - 2b_n} V_\pi^\star - V_{\widehat{\pi}}^\star \leq C\{e_n + a_n^{(1+\alpha)/\nu}\}.$$

Here empirical selection refers to maximization of the cross-fitted criterion over all feasible router–cutoff pairs, as specified for the theoretical search at the start of this appendix. For the selected route, the final cutoff maximizes $\widehat{F}^0$. The second uniform value bound, used twice, and the second bound in (29) therefore give

$$V_{\widehat{\pi}}^\star - F_\beta(\widetilde{g}) \leq C\{e_n + a_n^{(1+\alpha)/\nu}\}.$$

Adding these displays proves (18).

With fixed dimension, folds, and route class, the value fluctuations are $O_p(n^{-1/2})$. This remains their order when the budget buffer uses $b_n(\delta_n)$ with $\delta_n \downarrow 0$: use the cost event at confidence $1-\delta_n$ and the value bound at each fixed confidence level to establish the O_p statement.

A.3 Proof of Theorem 4

Fix a router π. Write $V = V_\pi^\star$ and $\tau = \tau_\pi = \tau_\pi^\star$, where the last notation is used in Section 4.3, and put $w = 1+\beta^2$, $H(d) = Pd + \beta^2\rho$. At its oracle value V, the rule $d^* = \mathbf{1}\{p_\pi > \tau\}$, with $\tau = V/w$, has zero Dinkelbach residual. Consequently

$$\begin{aligned} H(d)\{V - F_\beta(d)\} &= wP[p_\pi(d^* - d)] - VP(d^* - d) \\ &= wP[|p_\pi - \tau|\mathbf{1}\{d \neq d^*\}]. \end{aligned}$$

Taking $d = d_\pi^{\mathrm{raw}}$ proves (23).

We have the following auxiliary result.

Proposition 1 (Probability and cutoff error)**.** *Suppose $\rho \geq \rho_0 > 0$ and $P(0 < |p_\pi - \tau| \leq u) \leq Lu^\alpha$ for $0 < u \leq u_0$, where $L > 0$ and $\alpha \geq 0$. If $\|p^0 - p_\pi\|_\infty \leq a$, $|t - \tau| \leq h$, and $a + h \leq u_0$, then*

$$V - F_\beta(\mathbf{1}\{p^0 > t\}) \leq \frac{(1+\beta^2)L}{\beta^2\rho_0}(a+h)^{1+\alpha}.$$

Proof. Let $d = \mathbf{1}\{p^0 > t\}$. If $a+h = 0$, then $d = d^*$ almost surely and the claim follows. Otherwise, the bounds on p^0 and t imply that a disagreement with nonzero weight requires $0 < |p_\pi - \tau| \leq a+h$. Using the margin condition stated in this proposition with $u = a + h \leq u_0$, we obtain

$$\begin{aligned} P[|p_\pi - \tau|\mathbf{1}\{d \neq d^*\}] &\leq (a+h)P(0 < |p_\pi - \tau| \leq a+h) \\ &\leq L(a+h)^{1+\alpha}. \end{aligned}$$

Substituting this bound into the preceding regret identity and using $w = 1+\beta^2$ and $H(d) \geq \beta^2\rho_0$ gives the stated inequality. □

Let $\widehat{t}_{\mathrm{scr}}$ be the cutoff retained by the cross-fitted router search. Applied to each fold-specific fitted score, the proposition uses $h = |\widehat{t}_{\mathrm{scr}} - V/(1+\beta^2)|$; for the full-data score and final deployed cutoff, it uses $h = |\widehat{t} - V/(1+\beta^2)|$. Correct ranking alone need not make this cutoff error small.

Proof of Theorem 4. Since $a_n = o_p(1)$, the regret bound gives $T_n = V_n - F_\beta(\widetilde{g}) = o_p(1)$. Put $V_{\mathrm{raw},n} = \max_{\pi: C_\pi \leq B} F_\beta(d_\pi^{\mathrm{raw}})$, so $\Delta^{\mathrm{raw}} = V_n - V_{\mathrm{raw},n}$ by (24). On the simultaneous population-feasibility event, $F_\beta(\widehat{g}_{\mathrm{raw}}) \leq V_{\mathrm{raw},n}$, regardless of how the raw policy was selected. Hence

$$F_\beta(\widetilde{g}) - F_\beta(\widehat{g}_{\mathrm{raw}}) \geq \Delta^{\mathrm{raw}} - T_n.$$

Eventually $\Delta^{\mathrm{raw}} \geq 3\Delta_0/4$, while $P(T_n \leq \Delta_0/4) \to 1$, which proves the claim. A buffer $b = b_n(\delta_n)$ on the class-wide cost event ensures feasibility of both procedures; an unbuffered comparison needs its own feasibility argument. □

A.4 Proof of Theorem 2

Write $w = 1+\beta^2$, $b = \alpha/(1+\alpha)$, $k = (1+\alpha)/\nu$, and

$$q_n = n^{-(1+\alpha)/(2+\alpha)}, \qquad A_n = a_n^k = o_p(1).$$

The proof retains the score error a_n in Eq. (13), without requiring a particular convergence rate beyond $a_n = o_p(1)$. We first control the loss from selecting a router using the cross-fitted criterion. We then obtain a sharper bound for selecting the final cutoff after refitting on all observations, and add these two losses.

Throughout, Pf denotes expectation under the current population distribution P_n, and $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f_i$ its full-sample average, where f_i is f evaluated on observation i. When f is fitted, Pf evaluates the realized function on a fresh observation. All empirical bounds below hold uniformly over fixed classes of functions containing every possible fit; we therefore do not condition on the fitted functions.

Loss identity and an oracle-cutoff comparator. For an active router j, let $\tau = V_{j,n}/w$ be its optimal posterior cutoff, and suppress the router index in $d^* = \mathbf{1}\{p_j > \tau\}$, $S = |p_j - \tau|$, and

$$L(d) = wP\{S\mathbf{1}(d \neq d^*)\}.$$

Thus $L(d)$ is the probability of disagreeing with the oracle, weighted by the posterior distance S from its cutoff and scaled by w. The regret identity proved in Appendix A.3 and the margin condition in Assumption 3 give

$$V_{j,n} - F_\beta(d) = \frac{L(d)}{\beta^2\rho + Pd}, \qquad P(d \neq d^*) \leq CL(d)^b. \tag{30}$$

The denominator belongs to $[\beta^2\rho_0, w]$. The disagreement bound follows by splitting at $S = z$, obtaining $P(d \neq d^*) \leq Lz^\alpha + L(d)/(wz)$, and optimizing z. Constants may be enlarged outside the local margin range.

For router j, let $\bar{\tau}_j$ be its aligned population score cutoff and $\bar{S} = |\bar{p}_j - \bar{\tau}_j|$. Equation (28) gives $P\{S\mathbf{1}(\bar{S} \leq r)\} \leq Cr^k$. Thresholding a fitted score at $\bar{\tau}_j$ can disagree with the aligned population decision only where $\bar{S} \leq a_n$. Substituting $r = a_n$ therefore gives, simultaneously for every active router and every fold fit or full-data refit,

$$L\big(\mathbf{1}\{\widehat{p}_j^{(-v)} > \bar{\tau}_j\}\big) \leq CA_n. \tag{31}$$

Comparing routers using the cross-fitted criterion. For router j, let $\mathcal{D}_j$ contain all binary rules that, on each cell of the fixed partition χ, threshold an affine function of the fixed features $\Phi_{\pi_j(z)}(z, x)$ for the plan assigned to that cell. We allow strict and weak threshold inequalities and constant decisions. The link ψ in Eq. (10) is nondecreasing, so every rule $\mathbf{1}\{\widehat{p}_j^{(-v)} > t\}$ belongs to $\mathcal{D}_j$, regardless of the fitted coefficients and cutoff. This is the deterministic class used in the proof of Lemma 2: its VC dimension is bounded by a constant times KD, where $D = \max_a \dim \Phi_a$ as in Section 4.1. The class is fixed before observing the data even though a fitted rule selected from it is random.

Let I_v be fold v, put $\omega_v = |I_v|/n$, and write $\mathbb{P}_v f = |I_v|^{-1} \sum_{i \in I_v} f_i$ and $\widehat{\rho} = \mathbb{P}_n Y$. For a tuple $\mathbf{d} = (d_1, \ldots, d_V) \in \mathcal{D}_j^V$, define its empirical and population values by

$$\widehat{F}_j^-(\mathbf{d}) = \frac{w \sum_v \omega_v \mathbb{P}_v(Yd_v)}{\beta^2\widehat{\rho} + \sum_v \omega_v \mathbb{P}_v d_v},$$
$$\bar{F}_j(\mathbf{d}) = \frac{w \sum_v \omega_v P(Yd_v)}{\beta^2\rho + \sum_v \omega_v P d_v}.$$

For the algorithm's candidate (j, t), take $d_v = \mathbf{1}\{\widehat{p}_j^{(-v)} > t\}$. Then $\widehat{F}_j^-(\mathbf{d})$ is exactly the cross-fitted criterion in Eq. (8): each observation is evaluated using the fit that excludes its fold. The quantity $\bar{F}_j$ replaces the fold averages by population expectations of those same rules; it is the value of their mixture, not the value of the full-data refit.

The bounded-VC concentration argument in Lemma 2 gives, for each fixed fold,

$$\sup_{j,\, d \in \mathcal{D}_j} \big\{|(\mathbb{P}_v - P)d| + |(\mathbb{P}_v - P)(Yd)|\big\} = O_p(n^{-1/2}).$$

Here the number of routers and folds is fixed and $|I_v|$ is proportional to n. Weighting these bounds by ω_v controls the numerator and denominator above. Since $\widehat{\rho} - \rho = O_p(n^{-1/2})$ and $P(\widehat{\rho} \geq \rho_0/2) \to 1$, both denominators are bounded away from zero with probability tending to one. Subtracting the two ratios consequently yields

$$R_n := \sup_{j,\, \mathbf{d} \in \mathcal{D}_j^V} |\widehat{F}_j^-(\mathbf{d}) - \bar{F}_j(\mathbf{d})| = O_p(n^{-1/2}).$$

The bound is uniform over all tuples, so we may substitute the fitted rules without treating their overlapping training samples as independent. Moreover, $\bar{F}_j(\mathbf{d}) \leq V_{j,n}$ for every tuple: the definition of $V_{j,n}$ gives $wP(Yd_v) \leq V_{j,n}(\beta^2\rho + Pd_v)$ for each v, and summing with weights ω_v proves the claim.

To make the budget comparison explicit, let $\widehat{C}_j = n^{-1}\sum_i c_{i,\pi_j(Z_i)}$ be the empirical cost of router j, where π_j is its context-to-action map. Consider

$$\mathcal{C}_n = \left\{ \max_{j\in\mathcal{J}} |\widehat{C}_j - C_{j,n}| \le b_n(\delta_n) \right\}.$$

Hoeffding's inequality, applied to the bounded costs and then combined over the finite router class, gives $P(\mathcal{C}_n) \ge 1-\delta_n$. On this event, every active router is feasible under the algorithm's empirical cap for all sufficiently large n, since $2b_n(\delta_n) \le \kappa$ implies

$$\widehat{C}_j \le C_{j,n} + b_n(\delta_n) \le B - \kappa + b_n(\delta_n) \le B - b_n(\delta_n).$$

Conversely, any selected router satisfies $C_{\widehat{j},n} \le \widehat{C}_{\widehat{j}} + b_n(\delta_n) \le B$.

Choose a population-optimal router $j^\star$. It belongs to $\mathcal{J}_0$ because every feasible router outside $\mathcal{J}_0$ has a strictly positive gap. Let $\mathbf{d}^0$ be the tuple of its fold fits thresholded at $\bar{\tau}_{j^\star}$, and let $\widehat{\mathbf{d}}$ be the tuple selected by the cross-fitted search. Applying the loss identity to the fold mixture and using (31) gives $\bar{F}_{j^\star}(\mathbf{d}^0) \ge V_n - CA_n$. Let $\zeta_n \ge 0$ be the difference between the best feasible cross-fitted value and the selected value. For the exact theoretical search, $\zeta_n = 0$; the same argument allows $\zeta_n = o_p(n^{-1/2})$. Empirical selection on $\mathcal{C}_n$ gives

$$\begin{aligned} V_n - CA_n - R_n &\le \widehat{F}^-_{j^\star}(\mathbf{d}^0) \\ &\le \widehat{F}^-_{\widehat{j}}(\widehat{\mathbf{d}}) + \zeta_n \le V_{\widehat{j},n} + R_n + \zeta_n. \end{aligned}$$

Thus, with $Z_n = 2R_n + \zeta_n = O_p(n^{-1/2})$,

$$0 \le D_n := \Delta_{\widehat{j},n} \le CA_n + Z_n. \tag{32}$$

Since $A_n = o_p(1)$, this selected gap tends to zero in probability. Assumption 4 bounds every feasible gap outside $\mathcal{J}_0$ below by the fixed constant c, so $P(\widehat{j} \in \mathcal{J}_0) \to 1$. For those inactive routers, the comparison needed only the upper bound by their oracle values.

Controlling empirical error for decisions close to the oracle. Fix an active router j and use the class $\mathcal{D}_j$ just defined. For $d \in \mathcal{D}_j$, put $f_d = (wY - V_{j,n})(d - d^*)$. The loss identity implies $Pf_d = -L(d)$, so $(\mathbb{P}_n - P)f_d$ is the sampling error in comparing d with its oracle rule. We will bound this error more sharply when $L(d)$ is small, rather than use the global $O_p(n^{-1/2})$ bound.

We first record the empirical-process inequality used for this purpose. For a measurable set A with mass $p_A = P(A)$, consider functions supported on A, bounded in absolute value by w, and with second moments at most Cz, where $0 < z \le p_A$. For the restrictions to A of the classes used here, the local maximal inequality gives

$$\mathbb{E}^* \sup_f |(\mathbb{P}_n - P)f| \le C\left\{ \sqrt{\frac{z\log(ep_A/z)}{n}} + \frac{\log(ep_A/z)}{n} \right\}. \tag{33}$$

Here $\mathbb{E}^*$ denotes outer expectation, which agrees with expectation when the supremum is measurable. The factor p_A records that these functions vanish outside A: after restricting any covering measure to A and normalizing it, the same polynomial covering bound for $\mathcal{D}_j$ applies. Multiplication by the bounded function $wY - V_{j,n}$ and subtraction of the fixed oracle rule preserve that bound. Equation (33) then follows from Theorem 2.1 of van der Vaart and Wellner (2011). Without a second-moment restriction, the corresponding bounded-VC inequality gives $C\sqrt{p_A/n}$.

Fix a small constant $r_{\mathrm{loc}} > 0$. For $0 < r < r_{\mathrm{loc}}$, consider all decisions with $L(d) \le r$, and put $h = r^{1/(1+\alpha)}$. We partition observations according to their posterior distance $S = |p_j - \tau|$ from the oracle cutoff. The innermost band $\{S \le h\}$ has mass at most $Ch^\alpha = Cr^b$ by the posterior margin condition. Outside this band, use the sets $A_\ell = \{2^\ell h < S \le 2^{\ell+1}h\}$, $\ell = 0, 1, \ldots$. The same margin condition bounds their mass, while $L(d) = wP\{S\mathbf{1}(d \ne d^*)\} \le r$ bounds disagreement on each set:

$$P(A_\ell) \le Cr^b 2^{\alpha\ell}, \qquad \sup_{L(d)\le r} P\{(d-d^*)^2 \mathbf{1}_{A_\ell}\} \le \frac{r}{w2^\ell h} \le Cr^b 2^{-\ell}.$$

The second inequality reflects that an error farther from the cutoff costs more, so a decision with small $L(d)$ can make fewer such errors. Since $|f_d| \le w\mathbf{1}\{d \ne d^*\}$, it also bounds the second moment of $f_d\mathbf{1}_{A_\ell}$ up to a constant. When $P(A_\ell)$ exceeds the bound $Cr^b 2^{-\ell}$, apply (33) with z of that order. The ratio $P(A_\ell)/z$ is at most a constant times $2^{(1+\alpha)\ell}$, so its logarithm is $O(\ell+1)$. When the mass is smaller, use the bound $C\sqrt{P(A_\ell)/n}$ instead. Thus in either case,

$$\mathbb{E}^* \sup_{L(d)\le r} |(\mathbb{P}_n - P)(f_d\mathbf{1}_{A_\ell})| \le C\sqrt{\frac{r^b(\ell+1)2^{-\ell}}{n}} + \frac{C(\ell+1)}{n}.$$

The innermost band contributes at most $C\sqrt{r^b/n}$. There are $O(\log(e/r))$ nonempty outer bands because $S \le 1$. The series $\sum_{\ell\ge 0}\sqrt{(\ell+1)2^{-\ell}}$ converges, and the sum of the linear terms is $O(\log^2(e/r)/n)$. Adding these bounds yields

$$\mathbb{E}^* \sup_{d\in\mathcal{D}_j:\, L(d)\le r} |(\mathbb{P}_n - P)f_d| \le C\sqrt{r^b/n} + C\log^2(e/r)/n. \tag{34}$$

Extending the bound to all loss levels. For a fixed active router j, Eq. (34) bounds the expected supremum of $|(\mathbb{P}_n - P)f_d|$ over binary decision rules $d \in \mathcal{D}_j$ satisfying $L_j(d) \le r$. Here d maps the returned features U_j to a prediction in $\{0,1\}$, and $L_j(d)$ is its weighted disagreement loss defined above. We seek a bound that holds for every $j \in \mathcal{J}_0$ and every $d \in \mathcal{D}_j$, including the final fitted rule $\widehat{d}$, without restricting $L_j(d)$ to a prescribed radius. For any fixed $c_0 > 0$, write L_j and $f_{j,d}$ with their router indices explicit, and define

$$B_n = \max_{j\in\mathcal{J}_0}\sup_{d\in\mathcal{D}_j}\left[|(\mathbb{P}_n - P)f_{j,d}| - c_0 L_j(d)\right]_+, \qquad [x]_+ = \max\{x, 0\}.$$

By definition, for every $j \in \mathcal{J}_0$ and $d \in \mathcal{D}_j$,

$$|(\mathbb{P}_n - P)f_d| \le c_0 L(d) + B_n. \tag{35}$$

It remains to show $B_n = O_p(q_n)$. Substituting the radius $r = q_n$ into the right-hand side of Eq. (34) gives $C\sqrt{q_n^b/n} = Cq_n$ for its first term, while $C\log^2(e/q_n)/n = o(q_n)$ for its second term. Thus the expected empirical-error supremum over $\{d \in \mathcal{D}_j : L_j(d) \le q_n\}$ is $O(q_n)$; q_n is the upper bound on the losses in this class, not the loss of each individual rule.

Fix $T \ge 1$ and take n large enough that $Tq_n < r_{\mathrm{loc}}/2$. Partition the possible loss sizes into $L_j(d) \le Tq_n$, intervals

$$r_{m,n} < L_j(d) \le 2r_{m,n}, \qquad r_{m,n} = 2^m Tq_n, \quad m \ge 0, \quad r_{m,n} < r_{\mathrm{loc}}/2,$$

and the region $L_j(d) > r_{\mathrm{loc}}/2$; overlap with the last region is harmless. This grouping into intervals with endpoints doubling at each step is the dyadic peeling argument. Within each interval, the loss is bounded below by its left endpoint, while the local empirical bound can be applied at its right endpoint.

For $L_j(d) \le Tq_n$, an exceedance $B_n > Tq_n$ requires an empirical error greater than Tq_n. Markov's inequality and (34), applied at $r = Tq_n$, give

$$\mathbb{P}^*\left\{\max_{j\in\mathcal{J}_0}\sup_{d:\, L_j(d)\le Tq_n}|(\mathbb{P}_n - P)f_{j,d}| > Tq_n\right\} \le CT^{b/2-1} + \frac{C\log^2(e/q_n)}{nTq_n}.$$

We write $\mathbb{P}^*$ for outer probability to accommodate a possibly nonmeasurable supremum; it is ordinary probability when that supremum is measurable. The fixed number of active routers is absorbed into C. On the mth interval, an exceedance requires $|(\mathbb{P}_n - P)f_{j,d}| > c_0 r_{m,n}$. Applying Markov's inequality to the supremum at radius $2r_{m,n} < r_{\mathrm{loc}}$ bounds its probability by

$$\frac{C}{r_{m,n}}\left\{\sqrt{\frac{(2r_{m,n})^b}{n}} + \frac{\log^2(e/(2r_{m,n}))}{n}\right\} \le C(2^m T)^{b/2-1} + \frac{C\log^2(e/q_n)}{n\,2^m Tq_n}.$$

Both terms are summable over m: the first decreases geometrically because $b/2 - 1 < 0$, and the second contains 2^{-m}. For losses above $r_{\mathrm{loc}}/2$, an exceedance requires an empirical error greater than

the fixed constant $c_0 r_{\mathrm{loc}}/2$. The global bounded-VC inequality gives expected supremum $O(n^{-1/2})$, so Markov's inequality bounds this final probability by $Cn^{-1/2}$. A union bound over the loss intervals therefore gives

$$\mathbb{P}^*(B_n > Tq_n) \le CT^{b/2-1} + \frac{C\log^2(e/q_n)}{nTq_n} + Cn^{-1/2}.$$

Since $\alpha > 0$, we have $\log^2(e/q_n)/(nq_n) \to 0$. Taking the upper limit as $n \to \infty$ and then letting $T \to \infty$ proves that B_n/q_n is bounded in probability, as required.

Sharp loss at the final full-refit cutoff. Work on the event $\widehat{j} \in \mathcal{J}_0$. For the selected route, use the full-data comparator $d_0 = \mathbf{1}\{\widehat{p}_{\widehat{j}}^{(0)} > \bar{\tau}_{\widehat{j}}\}$. Equation (31) gives $L(d_0) \le CA_n$. For any binary rule d using this router's returned features, write

$$\widehat{F}_\beta(d) = \frac{w\mathbb{P}_n\{Yd(U_{\widehat{j}})\}}{\mathbb{P}_n d(U_{\widehat{j}}) + \beta^2\widehat{\rho}}.$$

Let $\widehat{d}$ be the final scalar-cutoff maximizer of this full-sample criterion for the refitted score, as in Algorithm 1, and put $v_0 = \widehat{F}_\beta(d_0)$. The full-sample uniform value bound in Lemma 2 gives $|v_0 - F_\beta(d_0)| = O_p(n^{-1/2})$. Combining it with the loss identity and $L(d_0) \le CA_n$ yields

$$\epsilon_n := |v_0 - V_{\widehat{j},n}| \le CA_n + O_p(n^{-1/2}).$$

On $\widehat{\rho} \ge \rho_0/2$, the empirical denominators are positive. The maximizing property $\widehat{F}_\beta(\widehat{d}) \ge v_0$ and the equality $\widehat{F}_\beta(d_0) = v_0$ imply

$$\mathbb{P}_n[(wY - v_0)\widehat{d}] \ge v_0\beta^2\widehat{\rho},$$
$$\mathbb{P}_n[(wY - v_0)d_0] = v_0\beta^2\widehat{\rho}.$$

Subtracting cancels the prevalence term and gives $\mathbb{P}_n[(wY - v_0)(\widehat{d} - d_0)] \ge 0$. Since $P[(wY - V_{\widehat{j},n})(\widehat{d} - d_0)] = L(d_0) - L(\widehat{d})$, adding and subtracting this population expectation and replacing v_0 by $V_{\widehat{j},n}$ gives

$$L(\widehat{d}) \le L(d_0) + \left|(\mathbb{P}_n - P)[(wY - V_{\widehat{j},n})(\widehat{d} - d_0)]\right| + \epsilon_n\mathbb{P}_n|\widehat{d} - d_0|.$$

Pointwise, $|\widehat{d} - d_0| \le \mathbf{1}\{\widehat{d} \ne d^*\} + \mathbf{1}\{d_0 \ne d^*\}$. Apply the disagreement bound in (30) to each term. The same uniform empirical bound for $\mathcal{D}_j$ and its pairwise disagreement class controls the difference between empirical and population probabilities, giving

$$\mathbb{P}_n|\widehat{d} - d_0| \le C\{L(\widehat{d})^b + A_n^b\} + O_p(n^{-1/2}).$$

Young's inequality, with $1/(1-b) = 1 + \alpha$, bounds the product by

$$\epsilon_n\mathbb{P}_n|\widehat{d} - d_0| \le \tfrac{1}{4}L(\widehat{d}) + CA_n + O_p(q_n)$$

with probability tending to one. The factor $L(\widehat{d})^b$ contributes at most $L(\widehat{d})/4 + C\epsilon_n^{1+\alpha}$, where $\epsilon_n^{1+\alpha} \le CA_n^{1+\alpha} + O_p(n^{-(1+\alpha)/2})$. For the product involving A_n^b, the term involving only A_n is $CA_n^{1+b} \le CA_n$ with probability tending to one, while Young's inequality bounds the mixed term $O_p(n^{-1/2})A_n^b$ by $CA_n + O_p(n^{-(1+\alpha)/2})$. Multiplying by the remaining $O_p(n^{-1/2})$ empirical error contributes at most $CA_n + O_p(n^{-1})$. Also $A_n^{1+\alpha} \le A_n$ with probability tending to one. These bounds require only $A_n = o_p(1)$, and $n^{-(1+\alpha)/2}$ and n^{-1} are both $o(q_n)$.

For the other empirical term, observe that $(wY - V_{\widehat{j},n})(\widehat{d} - d_0) = f_{\widehat{d}} - f_{d_0}$. Apply (35) separately to these two functions with $c_0 = 1/4$ and use the triangle inequality:

$$|(\mathbb{P}_n - P)(f_{\widehat{d}} - f_{d_0})| \le \tfrac{1}{4}\{L(\widehat{d}) + L(d_0)\} + 2B_n.$$

Combining the last two bounds with the empirical comparison gives $L(\widehat{d}) \le L(\widehat{d})/2 + CA_n + O_p(q_n)$. Moving $L(\widehat{d})/2$ to the left, and then using the denominator lower bound in (30), proves, for a constant C independent of n,

$$[L(\widehat{d}) - CA_n]_+ = O_p(q_n), \qquad [V_{\widehat{j},n} - F_\beta(\widetilde{g}) - CA_n]_+ = O_p(q_n). \tag{36}$$

Every fitted decision is covered by the same deterministic classes, so this argument permits the data-dependent route and every maximizing cutoff tie. It requires neither recovery of the population cutoff nor a comparison between the OOF and full-fit empirical objectives.

The total rate. If $0 < \gamma < 1/2$, then $Z_n = o_p(n^{-\gamma})$ in (32). On an event with probability tending to one, $Z_n \le c_\Delta n^{-\gamma}/2$ and the selected route is active. A positive selected gap is at least $c_\Delta n^{-\gamma}$, so $Z_n \le D_n/2$ and $D_n \le 2CA_n$; the same conclusion holds when $D_n = 0$. If $\gamma \ge 1/2$, active-route selection gives $D_n \le C_\Delta n^{-\gamma}$ with probability tending to one. Combining these bounds with (36) proves the stronger statement

$$[V_n - F_\beta(\widetilde{g}) - Ca_n^k]_+ = O_p\big(q_n + \mathbf{1}\{\gamma \ge 1/2\} n^{-\gamma}\big)\,.$$

On $\mathcal{C}_n$, the selected router is population feasible, so

$$0 \le V_n - F_\beta(\widetilde{g}) = \Delta_{\widehat{j},n} + \{V_{\widehat{j},n} - F_\beta(\widetilde{g})\}.$$

Both terms are nonnegative by their oracle definitions. The complement of this event has probability tending to zero and does not affect the stochastic order. This proves (20), with its random score-error term interpreted through the positive-part bound above. If additionally $a_n^k = o_p(n^{-\gamma})$ and $\gamma < 1/2$, (32) also implies $P(D_n = 0) \to 1$. The final cutoff changes no acquisition plan, so the screening cost guarantee is preserved.

B Boundary and route lower bounds

The proof of Theorem 3 separates three sources of regret. Assumption 6 collects the common conditions for the constructions below. Each construction gives a regret lower bound on a suitable instance with probability bounded away from zero. We obtain the boundary and route-selection lower bounds via Le Cam's two-point testing method: we construct two training distributions with small total variation distance but different optimal decisions, so a testing error forces regret. For score compression, we use a Gaussian limit for the scaled error in the fitted score direction and show that, with positive probability, every scalar cutoff of that score incurs regret. The argument proceeds as follows.

- **Uncertainty near the decision boundary (Appendix B.1).** We construct two distributions that are difficult to distinguish from training data but require different predictions near the posterior cutoff. A learner must therefore make errors near this boundary under at least one distribution, giving a lower bound of order $n^{-(1+\alpha)/(2+\alpha)}$. This bound allows arbitrary binary terminal rules.
- **Selecting between nearly tied routes (Appendix B.1).** We swap the quality of two routes whose oracle values differ by order $n^{-\gamma}$. When $\gamma \ge 1/2$, the data cannot reliably identify the better route. Selecting the worse route then incurs regret of order $n^{-\gamma}$, even if its terminal rule is optimal.
- **Information lost through a fitted score (Appendix C).** An irrelevant feature causes a perturbation of order $n^{-1/2}$ in the fitted score direction. No scalar cutoff can undo the resulting errors, yielding a lower bound of order $n^{-(1+\alpha)/(2\nu)}$ for rules that threshold the fitted score. We establish this construction for the class of smooth convex losses specified there.
- **Combining the bounds (end of Appendix C).** We realize all three constructions using one fixed loss, partition, feature map, and action class, and verify the main assumptions, including fitting regularity. The slowest applicable lower bound gives the worst-case exponent in Theorem 3; the route-selection term enters only when $\gamma \ge 1/2$.

Index the fixed router class as $\Pi = \{\pi_j : j \in \mathcal{J}\}$, and write $U_j = (Z, X_{\pi_j(Z)})$ for the information available for its terminal prediction. Thus $p_j = p_{\pi_j}$ and $V_{j,n} = V^\star_{\pi_j}$ in the notation of Section 4, with the subscript n displaying dependence on the distribution P_n. In both constructions below the context is constant, so choosing a router amounts to choosing one of two fixed call plans.

Assumption 6 (Fixed-class decision benchmark)**.** *For each sample size n, the development observations are independent with common distribution P_n and include the label and returned information of every route in a fixed finite class $\mathcal{J}$. Collection and deployment have the same conditional output and cost laws. At deployment, the action $\pi_j(Z)$ is chosen using pre-call context Z, before its returned prediction $X_{\pi_j(Z)}$ is observed; uncalled outputs cannot be used. Route costs lie in $[0, c_{\max}]$, and $\rho = P_n Y \geq \rho_0 > 0$. All population quantities below are evaluated under P_n. For each router $j \in \mathcal{J}$, let $p_j = P_n(Y = 1 \mid U_j)$, $V_{j,n} = \sup_d F_{\beta,P_n}\{d(U_j)\}$, $\tau_j = V_{j,n}/(1+\beta^2)$, and $d^\star_j = \mathbf{1}\{p_j > \tau_j\}$. The supremum defining $V_{j,n}$ is over all measurable binary rules d that use only the returned information U_j. For every $j \in \mathcal{J}$, the posterior margin satisfies*

$$P_n(|p_j - \tau_j| \leq z) \leq L_0 z^\alpha, \qquad 0 < z \leq z_0,$$

with fixed $\alpha, L_0, z_0 > 0$. Let $C_{j,n} = \mathbb{E}_{P_n}[C_{\pi_j(Z)}]$ be the population mean cost, $V_n = \max_{j:C_{j,n}\leq B} V_{j,n}$, and $\Delta_{j,n} = V_n - V_{j,n}$. A fixed nonempty active set $\mathcal{J}_0$ has cost slack $C_{j,n} \leq B - \kappa$ and satisfies

$$\Delta_{j,n} = 0 \quad \text{or} \quad c_\Delta n^{-\gamma} \leq \Delta_{j,n} \leq C_\Delta n^{-\gamma}.$$

Every feasible route outside $\mathcal{J}_0$ has gap at least a fixed $c_0 > 0$. For each active route, its oracle decision belongs to the fixed class $\mathcal{D}_j$ defined below. All constants, feature dimensions, and the numbers of routers and plans are fixed; for uniform statements all conditions hold from one common sample size n_0.

Assumption 6 specifies the population benchmark and the information available to a learner. The distributions P_n may vary with n, allowing the positive gaps $\Delta_{j,n}$ to shrink as the training sample grows.

For router j, let $a_{j,k}$ be the plan assigned to context cell k, so $\pi_j(Z) = a_{j,\chi(Z)}$. As in Appendix A.4, use the deterministic class of binary rules

$$\mathcal{D}_j = \left\{ d_\vartheta(U_j) = \mathbf{1}\left\{\Phi_{\pi_j(Z)}(Z, X_{\pi_j(Z)})^\top \vartheta_{\chi(Z)} > 0\right\} : \vartheta_k \in \mathbb{R}^{\dim \Phi_{a_{j,k}}},\ k = 1, \ldots, K \right\}.$$

Here Φ_a is the fixed finite-dimensional numerical feature map for plan a, including intercept terms, from (10). We also permit weak threshold inequalities and constant decisions. Coefficients are unrestricted and may differ across context cells, including cells assigned the same plan. This class has VC dimension bounded by a constant times KD, with $D = \max_a \dim \Phi_a$. The requirement $d^\star_j \in \mathcal{D}_j$ means that some choice of coefficients reproduces the oracle's binary predictions. It does not require the affine score to equal p_j. The learners in Theorem 5 may use arbitrary measurable terminal rules; $\mathcal{D}_j$ is used only to specify which oracle decisions the benchmark can represent.

The boundary construction will give exponent a_α, and the route-selection construction will additionally give exponent γ when $\gamma \geq 1/2$. Define their combined exponent by

$$a_\alpha = \frac{1+\alpha}{2+\alpha}, \qquad r_{\mathrm{dir}} = \begin{cases} a_\alpha, & 0 < \gamma < 1/2, \\ \min\{\gamma, a_\alpha\}, & \gamma \geq 1/2. \end{cases}$$

B.1 Information-theoretic lower bounds

For fixed $\alpha, \nu, \gamma > 0$ and $p > 1$, let $\mathfrak{P}_{\alpha,\nu,\gamma}$ be the class of all distribution sequences $\mathbf{P} = (P_n)_{n \geq n_0}$ satisfying Assumption 6 on a fixed observation space, with fixed routers and feature maps. The constants in the assumption are common across routers, sample sizes, and sequences in the class; they may depend on the fixed exponents and β.

The lower bound will hold even for the subclass with the following additional property. Under each P_n and for each active router j, there is a population least-squares score $\bar{p}_j$ with a cutoff $\bar{\tau}_j \in (0,1)$ such that $\mathbf{1}\{\bar{p}_j > \bar{\tau}_j\} = d^\star_j$ almost surely. The score distance $\bar{S}_j = |\bar{p}_j - \bar{\tau}_j|$ and posterior distance $S_j = |p_j - \tau_j|$ satisfy the inequalities below for $0 < r \leq r_F$ and $v \geq 1$, with $L_F, r_F > 0$ common

across sample sizes, active routers, and sequences:

$$\begin{aligned} P_n(\bar{S}_j \le r) &\le L_F r^{\alpha/\nu}, \\ P_n(\bar{S}_j \le r, S_j > v r^{1/\nu}) &\le L_F r^{\alpha/\nu} v^{-p}, \\ P_n(r < \bar{S}_j \le 2r, S_j \le r^{1/\nu}/v) &\le L_F r^{\alpha/\nu} v^{-p}. \end{aligned} \tag{37}$$

Both constructions below satisfy all three inequalities in (37). The first two are the score-margin conditions in (16); the third is an additional restriction satisfied by the lower-bound constructions. These least-squares scores serve as auxiliary population witnesses; the learner need not fit or threshold them. The proof at the end of Appendix C separately verifies the corresponding conditions for the population scores induced by its chosen fitting loss. We use $\mathfrak{P}_{\alpha,\nu,\gamma}$ for this restricted class below. A lower bound for it also applies to the larger class specified by Assumption 6 alone.

A learning algorithm $\mathcal{A}_n$ maps the n full-information training observations, and possibly an independent random seed, to a deployment policy. That policy selects a router and uses only its returned information for the terminal prediction; randomized selections and predictions are allowed. Conditional on the training sample and seed, $F_{\beta,P_n}(\mathcal{A}_n)$ denotes the population value of the resulting policy, with deployment randomization averaged into its confusion probabilities. In the probability below, P_n^n denotes the law of the n independent training observations from P_n and, for randomized algorithms, includes the independent training seed. The supremum ranges over sequences $\mathbf{P}$, using the nth distribution of each sequence at sample size n.

Theorem 5 (Universal boundary and route lower bounds)**.** *For every fixed $\alpha, \nu, \gamma > 0$ and $p > 1$, there exist a fixed finite router class $\mathcal{J}$, fixed feature maps, and common constants satisfying Assumption 6 and* (37)*, and constants $c, c_0 > 0$ such that, for all sufficiently large n,*

$$\inf_{\mathcal{A}_n} \sup_{\mathbf{P} \in \mathfrak{P}_{\alpha,\nu,\gamma}} P_n^n\big\{V_n - F_{\beta,P_n}(\mathcal{A}_n) \ge c n^{-r_{\mathrm{dir}}}\big\} \ge c_0. \tag{38}$$

The infimum is over all measurable learning algorithms just described. For every learner and every sufficiently large n, the proof supplies at least one of the constructed sequences in $\mathfrak{P}_{\alpha,\nu,\gamma}$ on which regret is at least $cn^{-r_{\mathrm{dir}}}$ with probability at least c_0.

Proof. Write $b = \beta^2$ and $w = 1 + b$. For a fixed router j, let $d(U_j)$ be a terminal rule and $d_j^\star = \mathbf{1}\{p_j > \tau_j\}$ its oracle rule. The regret identity from Appendix A.3 is

$$V_{j,n} - F_{\beta,P_n}(d) = \frac{w\mathbb{E}_{P_n}[|p_j - \tau_j||d - d_j^\star|]}{\mathbb{E}_{P_n} d + b\rho}. \tag{39}$$

Here d and $d_j^\star$ are evaluated at U_j. The identity also holds when $d(U_j) \in [0,1]$ is the conditional probability of predicting 1: since $d_j^\star$ is binary, $|d - d_j^\star|$ is then the conditional probability of disagreement. The denominator is at most $1 + b = w$, so

$$V_{j,n} - F_{\beta,P_n}(d) \ge \mathbb{E}_{P_n}[|p_j - \tau_j||d - d_j^\star|].$$

The first construction uses this inequality to turn errors in the terminal prediction into regret. The second constructs two routers with different oracle values, so selecting the suboptimal router incurs regret even when its terminal rule is optimal.

Boundary experiment. Fix a small $\delta > 0$, put $m = \delta^\alpha$ and $a = (1 - m)/2$, and let the returned feature $X \in \{-1, 0, 1\}$ have probabilities (a, m, a). The two distributions will differ only in the probability of $Y = 1$ at $X = 0$. Choose q_0 as the positive solution of

$$bmq_0^2 + a(1 + b)q_0 - a = 0, \qquad \text{equivalently} \qquad q_0 = \frac{a}{a + b(a + mq_0)}.$$

This choice has a specific purpose: if the posterior at $X = 0$ were q_0, while the posteriors at $X = -1, 1$ were $0, 1$, then predicting 1 only at $X = 1$ would have value wq_0 and posterior cutoff q_0. The middle atom would lie exactly at that cutoff; including it among the positive predictions would give the same value. As $\delta \to 0$, $q_0 \to 1/(1 + b)$, so q_0 stays a fixed positive distance from both 0 and 1.

For $\theta \in \{-1, 1\}$, define P_θ as the joint law of (Y, X) with the masses above and conditional probability $\eta_\theta(x) = P_\theta(Y = 1 \mid X = x)$ given by

$$\eta_\theta(-1) = 0, \qquad \eta_\theta(0) = q_0 + \theta\delta, \qquad \eta_\theta(1) = 1.$$

We suppress the dependence of P_θ on δ until choosing δ as a function of n. Its prevalence is $\rho_\theta = a + m(q_0 + \theta\delta)$. Consider the candidate rules $d_-(x) = \mathbf{1}\{x = 1\}$ and $d_+(x) = \mathbf{1}\{x \in \{0, 1\}\}$. Their respective F_β values under P_- and P_+, divided by w, are

$$\tau_- = \frac{a}{a + b\{a + m(q_0 - \delta)\}}, \qquad \tau_+ = \frac{a + m(q_0 + \delta)}{a + m + b\{a + m(q_0 + \delta)\}}.$$

To check that these are the oracle cutoffs, set $A = a + b(a + mq_0) = a/q_0$ and subtract q_0:

$$\tau_- - q_0 = \frac{bmq_0\delta}{A - bm\delta}, \qquad \tau_+ - q_0 = \frac{m(1 - bq_0)\delta}{A + m + bm\delta}.$$

Both differences are $O(m\delta)$. Hence, for sufficiently small δ, $q_0 - \delta < \tau_-$ and $q_0 + \delta > \tau_+$. At $X = -1, 1$, the posteriors $0, 1$ also lie on the indicated sides of each cutoff. Thus $d_\theta = \mathbf{1}\{\eta_\theta > \tau_\theta\}$; the threshold characterization of the fractional optimum verifies that d_θ is the oracle rule and its value is $V_\theta = w\tau_\theta$. Moreover, the posterior distance $S_\theta(x) = |\eta_\theta(x) - \tau_\theta|$ satisfies $\delta/2 \le S_\theta(0) \le 2\delta$, while $S_\theta(-1)$ and $S_\theta(1)$ are bounded below by a positive constant independent of δ and θ.

We next verify the score conditions in (37) for these distributions. With numerical features $(1, X)$, the population least-squares predictor before clipping is $\rho_\theta + X/2$: indeed, $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1 - m$, and $\mathbb{E}(XY) = a$. The corresponding score is $\bar{p}_\theta(x) = \min\{1, \max\{0, \rho_\theta + x/2\}\}$. Set $\bar{\tau}_\theta = \rho_\theta - \theta\delta^\nu$. For small δ, $\bar{\tau}_\theta \in (0, 1)$ and $\mathbf{1}\{\bar{p}_\theta > \bar{\tau}_\theta\} = d_\theta$ at all three values of X. For $\bar{S}_\theta = |\bar{p}_\theta - \bar{\tau}_\theta|$, we have $\bar{S}_\theta(0) = \delta^\nu$, and $\bar{S}_\theta(-1), \bar{S}_\theta(1)$ are bounded below by a positive constant.

Choose fixed $z_0, r_F > 0$ so that z_0 is below both posterior distances at $X = \pm 1$, and $2r_F$ is below both score distances there, uniformly for small δ and both alternatives. Every nonempty event in the local margin and score inequalities then consists only of $X = 0$, whose probability is $m = \delta^\alpha$. If $\{S_\theta \le z\}$ is nonempty for $z \le z_0$, then $z \ge \delta/2$, giving $m \le 2^\alpha z^\alpha$ and the posterior margin in Assumption 6. For the three score inequalities, the implications are as follows:

- If $\{\bar{S}_\theta \le r\}$ is nonempty, then $r \ge \delta^\nu$, so its probability m is at most $r^{\alpha/\nu}$.
- If $\{\bar{S}_\theta \le r, S_\theta > vr^{1/\nu}\}$ is nonempty, then $r \ge \delta^\nu$ and $vr^{1/\nu} < 2\delta$, hence $v < 2$. Its probability is therefore at most $2^p r^{\alpha/\nu} v^{-p}$.
- If $\{r < \bar{S}_\theta \le 2r, S_\theta \le r^{1/\nu}/v\}$ is nonempty, then $\delta^\nu/2 \le r < \delta^\nu$ and $\delta/2 \le r^{1/\nu}/v$, again implying $v < 2$. Its probability is at most $2^{\alpha/\nu+p} r^{\alpha/\nu} v^{-p}$.

Thus one fixed L_F verifies all three inequalities for every fixed $\nu > 0$ and $p > 1$.

It remains to show that training data cannot reliably distinguish P_+ from P_-. Write $\mathrm{kl}(u, v)$ for the Kullback–Leibler divergence from a Bernoulli law with parameter u to one with parameter v, and let P_θ^n denote the law of n independent observations from P_θ. The law of X is the same under both alternatives, and the label laws at $X = \pm 1$ agree. Only $X = 0$ contributes to the divergence:

$$\mathrm{KL}(P_+^n, P_-^n) = nm\ \mathrm{kl}(q_0 + \delta, q_0 - \delta) \le Cn\delta^{\alpha+2}.$$

The inequality follows from $\mathrm{kl}(u, v) \le (u - v)^2/\{v(1 - v)\}$, since $q_0 \pm \delta$ stay away from 0 and 1. Choose $\delta = c_* n^{-1/(\alpha+2)}$, where $c_* > 0$ is small enough that the displayed divergence is at most $1/8$. Pinsker's inequality then bounds the total variation distance between P_+^n and P_-^n by $1/4$. Consequently, every test $\widehat{\theta} \in \{-1, 1\}$ based on the training sample satisfies

$$\max_{\theta \in \{-1,1\}} P_\theta^n(\widehat{\theta} \ne \theta) \ge \frac{1 - \mathrm{TV}(P_+^n, P_-^n)}{2} \ge \frac{3}{8}.$$

To apply this testing bound to prediction regret, let $\widehat{d}(0) \in [0, 1]$ be the learned probability of predicting 1 at $X = 0$, conditional on the training sample and any training seed. Define $\widehat{\theta} = 1$ when $\widehat{d}(0) \ge 1/2$ and $\widehat{\theta} = -1$ otherwise. A testing error implies disagreement probability at least $1/2$ at $X = 0$, because the oracle predicts 1 there under P_+ and 0 under P_-. The mass $m = \delta^\alpha$ and distance $S_\theta(0) \ge \delta/2$ therefore give, by (39),

$$V_\theta - F_{\beta,\theta}(\widehat{d}) \ge m\frac{\delta}{2}\frac{1}{2} = \frac{\delta^{1+\alpha}}{4} \quad \text{whenever } \widehat{\theta} \ne \theta.$$

Here $F_{\beta,\theta}$ denotes population F_β under P_θ. Combining the testing and regret inequalities yields

$$\inf_{\widehat{d}} \max_{\theta\in\{-1,1\}} P_\theta^n\Big\{V_\theta - F_{\beta,\theta}(\widehat{d}) \ge \tfrac{1}{4}c_*^{1+\alpha}n^{-a_\alpha}\Big\} \ge \tfrac{3}{8}. \tag{40}$$

The infimum includes learned randomized prediction rules. An independent training seed has the same law under both alternatives, so adjoining it to the data leaves their divergence and total variation unchanged.

Route experiment. There are two equal-cost routes and constant pre-call context. Choose $0 < e_0 < 1/(2+\beta^2)$ and let $Y \sim \mathrm{Bernoulli}(1/2)$. Let B_1, B_2 be independent of each other and Y, with $B_j \sim \mathrm{Bernoulli}(e_j)$, and set $X_j = Y \operatorname{xor} B_j$, so X_j flips the label with probability e_j. For this construction, P_+ is the full-panel law of (Y, X_1, X_2) with $(e_1, e_2) = (e_0 - \Delta/2, e_0 + \Delta/2)$; P_- swaps these error probabilities. Training observes the full panel, whereas the terminal rule at deployment observes only the selected route's output. For either alternative and each route j, suppress the alternative index and write $V_j = V_{j,n}$. Then

$$p_j(x) = e_j + (1-2e_j)x, \qquad d_j^\star(x) = x, \qquad V_j = 1 - e_j.$$

To verify the oracle rule, its value divided by w is $\tau_j = (1-e_j)/(1+b)$. The choice of e_0 ensures $e_j < \tau_j < 1 - e_j$ for all sufficiently small Δ, so $d_j^\star(x) = \mathbf{1}\{p_j(x) > \tau_j\}$. Under P_+, route 1 is optimal; under P_-, route 2 is optimal. In either case the difference between their oracle values is exactly Δ.

Because p_j is affine in x, the population least-squares score is $\bar{p}_j = p_j$. Take $\bar{\tau}_j = \tau_j$. Then $\bar{S}_j = S_j = |p_j - \tau_j|$ is bounded below by a positive constant uniformly in small Δ, under both alternatives and both routes. Choose z_0 below this constant and $2r_F$ below it as well. The events in the posterior margin and all three inequalities of (37) are then empty for their stated local radii. This verifies these conditions for every fixed α, ν, p.

The transformation from (Y, X_1, X_2) to (Y, B_1, B_2) is invertible. The label law is unchanged, while the two independent Bernoulli error laws are swapped. Thus the divergence of the full training panels is

$$\mathrm{KL}(P_+^n, P_-^n) = n\{\mathrm{kl}(e_0 - \Delta/2, e_0 + \Delta/2) + \mathrm{kl}(e_0 + \Delta/2, e_0 - \Delta/2)\} \le Cn\Delta^2.$$

For $\gamma \ge 1/2$, set $\Delta = c_* n^{-\gamma}$ with a sufficiently small fixed $c_* > 0$. Then $n\Delta^2 \le c_*^2$, so we can again make the divergence at most $1/8$. The same Pinsker and testing inequalities give error probability at least $3/8$ under one of the two alternatives. A learner that selects a single route defines a test by declaring $+$ when it chooses route 1 and $-$ when it chooses route 2. A testing error selects the suboptimal route. Every terminal rule using that route has value at most its oracle value, so its regret relative to the optimal route is at least Δ.

For completeness, fix an alternative $\theta \in \{-1, 1\}$ and consider a learned policy that randomizes its route at deployment. Conditional on the training sample and seed, let ω_1, ω_2 be its route probabilities, with $\omega_1 + \omega_2 = 1$, and let $d_j(U_j) \in [0,1]$ be its conditional probability of predicting 1 after selecting route j. Define $F_j(d_j) = F_{\beta,P_\theta}\{d_j(U_j)\}$ and $H_j = \mathbb{E}_{P_\theta} d_j + b\rho$, where $\rho = 1/2$. Averaging the confusion probabilities over route selection gives

$$F_{\mathrm{mix}} = \frac{\sum_{j=1}^2 \omega_j H_j F_j(d_j)}{\sum_{j=1}^2 \omega_j H_j}, \qquad b/2 \le H_j \le w.$$

Let j_{bad} denote the suboptimal route under the current alternative, and let $V_{\mathrm{best}} = \max\{V_1, V_2\}$. Since $V_{\mathrm{best}} - F_{j_{\mathrm{bad}}}(d_{j_{\mathrm{bad}}}) \ge \Delta$ and both routes have value at most V_{best},

$$V_{\mathrm{best}} - F_{\mathrm{mix}} \ge \frac{\omega_{j_{\mathrm{bad}}} H_{j_{\mathrm{bad}}}}{\sum_j \omega_j H_j}\Delta \ge \frac{b}{4w}\Delta \quad \text{if } \omega_{j_{\mathrm{bad}}} \ge \frac{1}{2}.$$

Declare the testing alternative to be $+$ when $\omega_1 \ge 1/2$ and $-$ otherwise. A testing error assigns at least half the weight to the suboptimal route, so the last display and the testing bound also give regret of order $n^{-\gamma}$ with probability at least $3/8$ for randomized route policies, under one alternative. When $\gamma < 1/2$, $n\Delta^2 = c_*^2 n^{1-2\gamma}$ grows with n, so this construction does not give a testing-error bound bounded away from zero. Only the boundary construction is used in that case.

One fixed class and verification of the remaining assumptions. We now place both pairs of distributions on the same observation space with the same routers and numerical feature maps. Use two plans, each

with its own fixed prediction state, constant pre-call context, and equal known costs strictly below B. Label the two plans by $j = 1, 2$. For the constant router selecting plan j, let $\mathsf{X}_j = (Q_j, \zeta_j)$ be its full returned information, corresponding to X_a in Section 4, and take $\Phi_j(Z, \mathsf{X}_j) = (1, Q_j, \zeta_j)^\top$, where each sign ζ_j is uniform on $\{-1, 1\}$ and independent of (Y, Q_1, Q_2). In the boundary construction, set $Q_1 = Q_2 = X$ and $\zeta_1 = \zeta_2 = \zeta$, so the two routes return identical numerical evidence. In the route construction, set $Q_j = 2X_j - 1$ and take ζ_1, ζ_2 independent of each other and the full panel. The fixed feature map is the same in both constructions; only the joint distribution of its coordinates changes.

The signs are independent of (Y, Q_1, Q_2). Consequently, conditioning on them does not change the population posteriors, and their coefficients in the population least-squares regressions are zero. The numerical Gram matrices are $\mathrm{diag}(1, 1 - \delta^\alpha, 1)$ in the boundary construction and I_3 in the route construction, so their smallest eigenvalues are uniformly bounded away from zero. The oracle rules exhibited above threshold Q_j and have zero sign coefficient, so they belong to the fixed affine classes $\mathcal{D}_j$. These three-dimensional numerical features also give the representation used by the compression construction in Appendix C.

The added signs do not change the testing arguments. In the boundary construction, average the learned probability of predicting 1 at $X = 0$ over the sign and, if applicable, over the policy's route probabilities. This produces the scalar $\widehat{d}(0)$ used in the boundary test. The two routes are duplicates, so observing their full panel supplies no additional label information. In the route construction, the extra signs have a common independent law under P_+ and P_- and leave the divergence unchanged.

Collection and deployment use the same conditional output and cost laws. Each plan has its own feature map and probability model, so no pooling across plans is needed in these constructions. The prevalences are bounded below uniformly, and the fixed costs give common budget slack $\kappa > 0$. Set $\mathcal{J}_0 = \mathcal{J}$. In the boundary construction both routers are optimal and have gap zero; in the route construction the optimal router has gap zero and the other has gap $c_* n^{-\gamma}$. Thus the active-gap condition holds, and there are no inactive routers to check. The posterior and score conditions were verified above with constants independent of n. Taking the larger of the required upper-bound constants and the smaller of the positive lower-bound constants makes them common to the two constructions.

Finally, for $0 < \gamma < 1/2$, the boundary bound (40) gives exponent $r_{\mathrm{dir}} = a_\alpha$. For $\gamma \geq 1/2$, the class contains both the boundary and route alternatives. For every learner, its worst-case regret over this class is therefore bounded below at the slower of n^{-a_α} and $n^{-\gamma}$, namely $n^{-\min\{a_\alpha, \gamma\}}$, with a fixed positive probability. This proves (38). □

C Proof of Theorem 3

The remaining lower-bound term in Theorem 3 comes from estimating a compressed score. Theorem 6 below first establishes this term for any one fixed loss satisfying the smoothness, penalty, and existence conditions stated here. At the end of the appendix, we choose one particular admissible loss and verify the main assumptions on all three constructions to complete the proof of Theorem 3.

We obtain the score-compression lower bound via a Gaussian limit for the scaled coefficient of an irrelevant feature. We first construct a distribution whose population fit assigns zero coefficient to that feature. A projection of the Gaussian limit then yields an event of positive probability on which, for every auxiliary empirical fit, the ratio of the irrelevant-feature coefficient to the informative-feature coefficient is at least a fixed multiple of $n^{-1/2}$. Finally, we show that this coefficient ratio forces every scalar cutoff to disagree with the optimal decision, and integrate that disagreement to obtain regret of order $n^{-(1+\alpha)/(2\nu)}$.

Fix one fitting loss, including its regularization, and one partition. Label the single fixed plan a_0, and write $W \in \mathbb{R}^{d_W}$ for its numerical predictors, excluding the intercept. In the constructions below, $(1, W^\top)^\top$ is the feature vector $\Phi_{a_0}(Z, X_{a_0})$ from (10); thus this notation describes a particular choice of the plan-specific model in Section 4. The empirical average is $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(Y_i, W_i)$. The

loss and its full-data fit have the form

$$\ell(y, W; (a, v)) = L_y(a + v^\top W) + \mathrm{pen}(v),$$
$$(\widehat{a}, \widehat{v}) \in \arg\min_{a\in\mathbb{R},\, v\in\mathbb{R}^{d_W}} \mathbb{P}_n \ell(Y, W; (a, v)). \tag{41}$$

Here a is the intercept, while a_0 is the plan index, and v is the slope vector. In the notation of Appendix A.1, $\theta_{a_0} = (a, v)$, $\ell_{a_0}(\theta) = \ell(Y, W; \theta)$, and $\widehat{\theta}_{a_0} = (\widehat{a}, \widehat{v})$. The fixed penalty pen is finite, convex, continuously differentiable, and radial, meaning that it depends only on $\|v\|_2$. The intercept is unpenalized. This includes $\mathrm{pen} \equiv 0$, ridge penalties, and smooth nonquadratic penalties on the slope norm. After fitting, the learner may select a router and a scalar cutoff using the same data. Thresholding a continuous strictly increasing link of $\widehat{a} + \widehat{v}^\top W$ gives the same family of binary rules as thresholding this affine score; clipping restricts that family. A lower bound valid for every affine-score cutoff therefore also covers these linked or clipped scores.

To analyze the fixed-penalty fit, introduce auxiliary fits with quadratic penalties:

$$(\widehat{a}_\lambda, \widehat{v}_\lambda) \in \arg\min_{a,v} \left\{ \mathbb{P}_n L_Y(a + v^\top W) + \frac{\lambda}{2}\|v\|_2^2 \right\}, \qquad \lambda \geq 0. \tag{42}$$

Their population counterparts minimize

$$Q_\lambda(a, v) = \mathbb{E}[L_Y(a + v^\top W)] + \frac{\lambda}{2}\|v\|_2^2; \qquad (a_\lambda, v_\lambda) \in \arg\min_{a,v} Q_\lambda(a, v).$$

The parameter λ indexes a family used in the proof. We will show that the fit with the prescribed penalty pen equals one of these auxiliary fits at a possibly data-dependent value $\widehat{\lambda}$. We therefore need a lower-bound event that holds simultaneously for all auxiliary fits, so that it also holds at $\widehat{\lambda}$.

Assume each label-specific base loss $L_y : \mathbb{R} \to \mathbb{R}$ is finite, strictly convex, and C^2. On the distribution constructed below, assume that every Q_λ, $\lambda \geq 0$, and the constant-score objective $a \mapsto \mathbb{E}L_Y(a)$ attain their minima, with all minimizers in one deterministic compact set. Embed a constant-score minimizer as $(a, 0)$ when describing this common parameter set. With probability tending to one, all corresponding empirical objectives must also attain their minima, and one deterministic compact set must contain their minimizers simultaneously over $\lambda \geq 0$. These are unconstrained minimizers; any imposed coefficient constraints are required to be inactive on that event. On a compact score interval containing these population and empirical scores, assume

$$\Delta(s) := L_1'(s) - L_0'(s) < 0. \tag{43}$$

Thus increasing the score has a smaller loss derivative for label 1 than for label 0. The constants and compact sets may depend on the fixed loss. We verify the stated existence and compactness conditions for squared, logistic, and exponential losses below. On the exceptional event where an auxiliary empirical minimizer is not finite or the empirical design lacks full rank, define the fitted rule arbitrarily; the probability of this event tends to zero. For the following construction, specialize to $d_W = 2$ numerical predictors, so the augmented design including the intercept has three coordinates.

Theorem 6 (Compression lower bound for a fixed convex loss)**.** *Fix $\alpha, \nu, \beta > 0$ and one fitting loss ℓ satisfying the preceding conditions. There exists a distribution P, independent of n, with constant pre-call context, a fixed one-cell partition, bounded three-dimensional augmented features, and one route whose mean cost is at most $B - \kappa$ for some $\kappa > 0$. The constant sequence $P_n = P$ satisfies Assumption 6 for every fixed $\gamma > 0$ and* (37) *for every fixed $p > 1$. Write V for this route's oracle F_β value, and define the binary rule obtained by thresholding the fitted affine score:*

$$\widehat{d}_u(W) = \mathbf{1}\{\widehat{a} + \widehat{v}^\top W > u\}.$$

There are $c_1, \epsilon > 0$, independent of n, such that

$$\liminf_{n\to\infty} \mathbb{P}\left\{ \inf_{u\in\mathbb{R}} \left[V - F_\beta(\widehat{d}_u) \right] \geq c_1 n^{-(1+\alpha)/(2\nu)} \right\} \geq \epsilon. \tag{44}$$

The probability in (44) *is over the training sample. For each realized fit, $F_\beta(\widehat{d}_u)$ evaluates its population performance on a fresh observation. The infimum over u means that the lower-bound event applies to every cutoff of that same fit, including a cutoff selected from data.*

The population least-squares score is decision-aligned. For that score, its distance $\bar{S}$ from an aligned cutoff and the posterior distance S satisfy $\bar{S} = CS^\nu$ for a constant $C > 0$. The prescribed fitting loss also has a decision-aligned population affine score on this distribution. Thus the lower bound comes from sampling error in the fitted coefficients, despite the existence of an aligned population score.

Proof. We construct a feature X that determines the oracle decision and an independent feature ζ that carries no label information. We then show that, with probability bounded away from zero, every auxiliary fit has an error of order at least $n^{-1/2}$ in the ratio of the ζ coefficient to the X coefficient. For such a fitted score, changing its scalar cutoff cannot recover the oracle decisions for both signs of ζ. Finally, representing the prescribed fixed-penalty fit as one of the auxiliary fits transfers the bound to (41).

A common hard distribution. Let T be symmetric on $[-1, 1]$, with $\mathbb{P}(|T| \le t) = t^\alpha$ for $0 \le t \le 1$. Let ζ be an independent uniform sign and set $W = (X, \zeta)$. Define

$$X = \operatorname{sgn}(T)|T|^\nu, \qquad \mu_T = \mathbb{E}|T| = \frac{\alpha}{\alpha+1}, \quad w = 1 + \beta^2, \qquad \tau = \frac{\mu_T}{4w}, \quad c = \frac{\beta^2 \mu_T}{8w^2}.$$

Because $T = \operatorname{sgn}(X)|X|^{1/\nu}$, it is determined by W. Draw $Y \mid W$ as Bernoulli with mean $\eta(W) = \tau + cT$. This is the true conditional probability $p^\star_{a_0}(Z, X_{a_0})$ for the single plan in this construction. The inequalities $0 < 2c < \tau < 1/4$ imply $\eta(W) \in [\tau - c, \tau + c] \subset (0, 1)$, so both label probabilities are bounded away from zero. Symmetry of T gives $\rho = \mathbb{E}Y = \tau$, and the definitions of c and τ give

$$\mathbb{E}(\eta - \tau)_+ = \frac{c\mu_T}{2} = \beta^2\tau^2.$$

The equation $\mathbb{E}(\eta - \tau)_+ = \beta^2\rho\tau$ is the zero fractional-objective residual at value $w\tau$. It identifies the optimal rule as $d^* = \mathbf{1}\{\eta > \tau\} = \mathbf{1}\{X > 0\}$ and its value as $V = w\tau$. The augmented design has Gram matrix

$$\mathbb{E}[(1, X, \zeta)^\top (1, X, \zeta)] = \operatorname{diag}\left(1, \frac{\alpha}{\alpha + 2\nu}, 1\right).$$

For comparison with the score assumptions, the population least-squares score on these features is

$$\bar{p}_{\mathrm{LS}}(W) = \tau + b_{\mathrm{LS}} X, \qquad b_{\mathrm{LS}} = \frac{\mathbb{E}(XY)}{\mathbb{E}X^2} = c\frac{\alpha + 2\nu}{\alpha + \nu + 1}.$$

Because $0 < b_{\mathrm{LS}} < 2c < \tau$ and $\tau + b_{\mathrm{LS}} < 1$, clipping leaves $\bar{p}_{\mathrm{LS}}$ unchanged. At cutoff τ, its predictions equal d^*. For $S = |\eta - \tau|$ and $\bar{S} = |\bar{p}_{\mathrm{LS}} - \tau|$, this gives

$$S = c|T|, \qquad \bar{S} = b_{\mathrm{LS}}|T|^\nu, \qquad \bar{S} = \frac{b_{\mathrm{LS}}}{c^\nu} S^\nu.$$

In particular $\mathbb{P}(S \le x) = (x/c)^\alpha$ for $0 < x \le c$. For completeness, put $q = \alpha/\nu$ and $A = c\, b_{\mathrm{LS}}^{-1/\nu}$, so that $S = A\bar{S}^{1/\nu}$. For $0 < r \le b_{\mathrm{LS}}/2$,

$$\mathbb{P}(\bar{S} \le r) = b_{\mathrm{LS}}^{-q} r^q.$$

On $\{\bar{S} \le r\}$, the relation $S = A\bar{S}^{1/\nu}$ implies $S \le Ar^{1/\nu}$. Thus the event $\{\bar{S} \le r, S > vr^{1/\nu}\}$ in the second inequality of (37) is empty when $v \ge A$. When $1 \le v < A$, its probability is at most $b_{\mathrm{LS}}^{-q} r^q$, which is at most $b_{\mathrm{LS}}^{-q} A^p r^q v^{-p}$. Similarly, on $\{r < \bar{S} \le 2r\}$ we have $S > Ar^{1/\nu}$. The third event $\{r < \bar{S} \le 2r, S \le r^{1/\nu}/v\}$ is therefore empty when $v \ge A^{-1}$; otherwise its probability is bounded by $\mathbb{P}(\bar{S} \le 2r) \le 2^q b_{\mathrm{LS}}^{-q} r^q$ and $v^p \le A^{-p}$. Together with the displayed formula for $\mathbb{P}(\bar{S} \le r)$, these bounds verify all three inequalities for any fixed $p > 1$, for example with

$$r_F = b_{\mathrm{LS}}/2, \qquad L_F = 2^q b_{\mathrm{LS}}^{-q} \max\{1, A^p, A^{-p}\}.$$

For Theorem 6, use one plan at a deterministic cost below B, constant pre-call context, and one returned prediction containing both coordinates of W. Collection and deployment use the same law, and the plan-specific probability model uses both coordinates. The only router is optimal and has gap zero, as permitted for every $\gamma > 0$. The margin calculation, the affine oracle rule $\mathbf{1}\{X > 0\}$, and the fixed cost slack verify Assumption 6. For the later proof of Theorem 3, make two copies of this plan with identical returned vectors, including the same ζ. The fixed partition is $\chi(Z) \equiv 1$, and each of the two routers selects its corresponding plan. Their numerical training samples, and hence their unique fitted scores, coincide.

Population coefficients of the auxiliary fits. Write $\dot{L}_y = L'_y$ and $v = (b, g)$, where b multiplies the informative feature X and g multiplies the independent sign ζ. For every finite $\lambda \geq 0$, the population objective is

$$Q_\lambda(a, b, g) = \mathbb{E}L_Y(a + bX + g\zeta) + \frac{\lambda}{2}(b^2 + g^2).$$

For fixed (a, b) and $g \neq 0$, averaging over the two equally likely signs of ζ and applying strict convexity gives

$$\frac{L_y(a + bX + g) + L_y(a + bX - g)}{2} > L_y(a + bX).$$

Since ζ is independent of (X, Y) and the quadratic penalty also weakly decreases when g is set to zero, the population minimizer has $g_\lambda = 0$.

Let a_{const} be the unique minimizer of the constant-score objective $a \mapsto \mathbb{E}L_Y(a)$. It satisfies

$$(1 - \tau)\dot{L}_0(a_{\text{const}}) + \tau\dot{L}_1(a_{\text{const}}) = 0.$$

Put $p_0 = \mathbb{E}[X\eta(W)] = c\mathbb{E}|T|^{\nu+1} > 0$. Using $\mathbb{E}X = 0$ and $\mathbb{E}[\dot{L}_Y(s) \mid W] = \dot{L}_0(s) + \eta(W)\Delta(s)$, we obtain

$$\left.\frac{\partial Q_\lambda}{\partial b}\right|_{(a,b,g)=(a_{\text{const}},0,0)} = \mathbb{E}[X\dot{L}_Y(a_{\text{const}})] = \Delta(a_{\text{const}})p_0 < 0.$$

Minimizing over the intercept defines the convex function $b \mapsto \inf_a Q_\lambda(a, b, 0)$. Its derivative at $b = 0$ is the displayed negative number, because the minimizing intercept there is uniquely a_{const}. This function therefore attains its minimum at $b_\lambda > 0$. Consequently the population affine score $a_\lambda + b_\lambda X$ gives the oracle rule d^* at cutoff a_λ. We later transfer this conclusion to the prescribed radial penalty by representing its minimizer as an auxiliary one.

Rescaling the slopes and proving uniform consistency. The parameter λ ranges over an unbounded interval. To obtain one bound for all auxiliary fits, write $t = (1 + \lambda)^{-1} \in (0, 1]$ and rescale the slopes as $b = tk$ and $g = tr$. Here r is the rescaled coefficient of ζ. As $\lambda \to \infty$, $t \to 0$; $t = 1$ corresponds to $\lambda = 0$. For $\lambda(t) = (1 - t)/t$, define

$$\widehat{\theta}_t = (\widehat{a}_t, \widehat{k}_t, \widehat{r}_t) := \left(\widehat{a}_{\lambda(t)}, \widehat{b}_{\lambda(t)}/t, \widehat{g}_{\lambda(t)}/t\right), \qquad t > 0.$$

Substituting this rescaling into the auxiliary first-order equations uses $\lambda b = (1-t)k$ and $\lambda g = (1-t)r$, giving

$$\begin{aligned} \mathbb{P}_n\dot{L}_Y(a + t(kX + r\zeta)) &= 0, \\ \mathbb{P}_n X\dot{L}_Y(a + t(kX + r\zeta)) + (1 - t)k &= 0, \\ \mathbb{P}_n\zeta\dot{L}_Y(a + t(kX + r\zeta)) + (1 - t)r &= 0. \end{aligned} \tag{45}$$

For $\theta = (a, k, r)$, let $\widehat{F}_t(\theta)$ be the vector of these three left-hand sides, and let $F_t(\theta)$ replace each $\mathbb{P}_n$ by its population expectation. These functions remain defined and continuous at $t = 0$. The population equations there have the unique solution

$$\theta_0 = (a_{\text{const}}, k_0, 0), \qquad k_0 = -\Delta(a_{\text{const}})p_0 > 0.$$

For the empirical endpoint, let $\widehat{a}_{\text{const}}$ minimize $\mathbb{P}_n L_Y(a)$ and set

$$\widehat{\theta}_0 = \big(\widehat{a}_{\text{const}}, -\mathbb{P}_n X\dot{L}_Y(\widehat{a}_{\text{const}}), -\mathbb{P}_n\zeta\dot{L}_Y(\widehat{a}_{\text{const}})\big).$$

These are exactly the solutions of (45) at $t = 0$; their last two coordinates are scaled slopes, not the zero slopes of a constant score.

All population roots lie in one compact set of (a, k, r) values. For $t \geq 1/2$, this follows by dividing the assumed bounded slopes (b, g) by $t \geq 1/2$. For $t \leq 1/2$, the last two equations express k and r as bounded loss-gradient averages divided by $1 - t \geq 1/2$. The same reasoning bounds all empirical roots on the assumed event of uniformly bounded empirical fits. For $t > 0$, strict convexity and the positive definite population Gram matrix give a unique population root, denoted $\theta_t = (a_t, k_t, 0)$; the endpoint root is also unique.

The function $t \mapsto \theta_t$ is continuous on $[0, 1]$. Indeed, if $t_m \to t$, compactness gives a convergent subsequence of θ_{t_m}; continuity of F_t makes its limit a root of F_t, and uniqueness identifies that limit

as θ_t. For $t > 0$, $k_t = b_{\lambda(t)}/t > 0$, while $k_0 > 0$. Continuity on the compact interval thus gives $\inf_t k_t > 0$.

Let $\mathcal{K}$ be a common compact set containing the population roots and, with probability tending to one, all empirical roots. For every $\varepsilon > 0$, uniqueness and continuity imply

$$\inf_{\substack{0\le t\le 1,\ \theta\in\mathcal{K} \\ \|\theta-\theta_t\|\ge\varepsilon}} \|F_t(\theta)\| > 0.$$

Thus parameters a fixed distance from the population root cannot nearly solve the population equations. The bounded, continuous finite-dimensional class defining these equations satisfies

$$\sup_{t\in[0,1],\,\theta\in\mathcal{K}} \|\widehat{F}_t(\theta) - F_t(\theta)\| = o_p(1).$$

Since $\widehat{F}_t(\widehat{\theta}_t) = 0$, the last two displays force every empirical root to approach its population counterpart:

$$\sup_{0\le t\le 1} \|\widehat{\theta}_t - \theta_t\| = o_p(1). \tag{46}$$

Expanding the fitted coefficient of the irrelevant feature. Write $O = (Y, X, \zeta)$, let $Pf = \mathbb{E}[f(O)]$, and define $m_{a,b}(O) = \zeta \dot{L}_Y(a + bX)$. Independence of ζ from (X, Y) and $\mathbb{E}\zeta = 0$ give

$$Pm_{a,b} = \mathbb{E}[\zeta \dot{L}_Y(a + bX)] = 0 \qquad \text{for every } (a, b). \tag{47}$$

We apply this identity to the third equation in (45), which determines the rescaled coefficient $\widehat{r}_t$ of ζ. Holding $(\widehat{a}_t, \widehat{k}_t)$ fixed and expanding only in $\widehat{r}_t$ gives the exact equality

$$0 = \mathbb{P}_n m_{\widehat{a}_t, t\widehat{k}_t} + \widehat{r}_t \widehat{H}_t, \qquad \widehat{H}_t = 1 - t + t\mathbb{P}_n \int_0^1 L''_Y(\widehat{a}_t + t\widehat{k}_t X + vt\widehat{r}_t\zeta)\, dv.$$

The factor $\zeta^2 = 1$ removes the sign from the second term. Uniform continuity of L''_y on compact score intervals, (46), and a uniform law of large numbers imply

$$\sup_{t\in[0,1]} |\widehat{H}_t - H_t| = o_p(1), \qquad H_t = 1 - t + t\mathbb{E}L''_Y(a_t + tk_t X).$$

The function $t \mapsto H_t$ is continuous and positive. For $t < 1$, $H_t \ge 1 - t > 0$. At $t = 1$, the positive slope k_1 makes $a_1 + k_1 X$ range over a nontrivial interval with full support. Strict convexity and C^2 smoothness imply $L''_y \ge 0$ and prevent it from vanishing throughout that interval. Since both labels have conditional probabilities bounded away from zero, the average curvature H_1 is positive. Continuity on $[0, 1]$ therefore gives $\inf_t H_t > 0$.

Define the centered empirical process $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f$. Because $Pm_{a,b} = 0$ for every pair (a, b), the fitted equation becomes

$$\sqrt{n}\,\widehat{r}_t = -\frac{\mathbb{G}_n m_{\widehat{a}_t, t\widehat{k}_t}}{\widehat{H}_t}.$$

For (a, b) in the compact coefficient set used here, the functions $m_{a,b}$ are bounded and Lipschitz in their parameters, since L''_y is bounded on the corresponding score interval. The empirical process indexed by this finite-dimensional class has a functional Gaussian limit and is stochastically equicontinuous. In particular, uniform consistency of the fitted parameters gives

$$\sup_t \left|\mathbb{G}_n\{m_{\widehat{a}_t, t\widehat{k}_t} - m_{a_t, tk_t}\}\right| = o_p(1), \qquad \sup_t |\mathbb{G}_n m_{a_t, tk_t}| = O_p(1).$$

Thus replacing the fitted intercept and signal slope inside this centered process by their population values changes it by $o_p(1)$, uniformly over t. Together with the uniform convergence of $\widehat{H}_t$ and $\widehat{k}_t$ to positive limits, this yields, for $D_t = k_t H_t$,

$$\sup_{0\le t\le 1} \left|\sqrt{n}\frac{\widehat{r}_t}{\widehat{k}_t} - \frac{1}{\sqrt{n}}\sum_{i=1}^n f_t(O_i)\right| = o_p(1), \qquad f_t(O) = -\frac{\zeta \dot{L}_Y(a_t + tk_t X)}{D_t}. \tag{48}$$

Here D_t is continuous and bounded above and below by positive constants, and $Pf_t = 0$ by (47). For $t > 0$, the coefficient ratio in (48) is $\widehat{r}_t/\widehat{k}_t = \widehat{g}_{\lambda(t)}/\widehat{b}_{\lambda(t)}$. It measures the fitted coefficient of the irrelevant feature relative to that of the informative feature. At $t = 0$, it uses the scaled endpoint coefficients defined above. The process $\mathbb{Z}_n(t) = n^{-1/2}\sum_i f_t(O_i)$ converges in distribution, with respect to the supremum norm on $[0, 1]$, to a mean-zero Gaussian process G_t with continuous sample paths and covariance $\mathrm{Cov}(G_s, G_t) = P(f_s f_t)$.

A positive fluctuation shared by all auxiliary fits. Define $\xi(O) = \zeta\{Y - \eta(W)\}$ and $\Xi_n = n^{-1/2}\sum_i \xi(O_i)$. The pair $(\mathbb{Z}_n, \Xi_n)$ converges jointly to (G, Ξ), where Ξ is centered Gaussian with variance $\sigma_\xi^2 = P[\eta(1-\eta)] > 0$. Conditional expectation over Y gives

$$\mathbb{E}[\dot{L}_Y(s)(Y-\eta) \mid W] = \eta(1-\eta)\{\dot{L}_1(s) - \dot{L}_0(s)\} = \eta(1-\eta)\Delta(s).$$

Applying this at $s = a_t + tk_tX$ in $P(f_t\xi)$ yields

$$\mathrm{Cov}(G_t, \Xi) = -\frac{P[\eta(1-\eta)\Delta(a_t + tk_tX)]}{D_t} > 0.$$

The covariance is continuous in t, and its minimum on $[0,1]$ is positive. Define the Gaussian projection coefficients and residuals by

$$c_t = \frac{\mathrm{Cov}(G_t, \Xi)}{\sigma_\xi^2}, \qquad c_{\min} = \min_{t\in[0,1]} c_t > 0, \qquad \mathcal{R}_t = G_t - c_t\Xi.$$

Each residual has zero covariance with Ξ. Joint Gaussianity therefore makes the entire residual process independent of Ξ. Its continuous paths satisfy $\sup_t |\mathcal{R}_t| < \infty$ almost surely. Choose $B_R < \infty$ so that $\mathbb{P}(\sup_t |\mathcal{R}_t| \le B_R) > 0$, and fix $\delta > 0$. The event

$$\left\{\sup_t |\mathcal{R}_t| \le B_R\right\} \cap \left\{\Xi > \frac{B_R + 2\delta}{c_{\min}}\right\}$$

has positive probability by independence and the nondegenerate normal tail of Ξ. On this event, $G_t = c_t\Xi + \mathcal{R}_t > 2\delta$ for every t. Hence $\mathbb{P}(\inf_t G_t > 2\delta) > 0$.

The infimum over $t \in [0,1]$ is a continuous functional in the supremum norm. Weak convergence of $\mathbb{Z}_n$ and the uniform remainder in (48) thus imply

$$\liminf_n \mathbb{P}\left\{\inf_{0\le t\le 1} \sqrt{n}\frac{\widehat{r}_t}{\widehat{k}_t} > \delta\right\} > 0. \tag{49}$$

Also $\inf_t \widehat{k}_t > 0$ with probability tending to one, by (46) and $\inf_t k_t > 0$. Intersecting these two events preserves the positive lower limit of their probability. On their intersection, every finite auxiliary penalty has $\widehat{b}_\lambda > 0$ and $\widehat{g}_\lambda/\widehat{b}_\lambda > \delta/\sqrt{n}$. Because this statement is simultaneous over λ, it also holds at the data-dependent value $\widehat{\lambda}$ representing the prescribed fixed-penalty fit.

Why every scalar cutoff incurs regret. Fix an affine score $a + bX + g\zeta$ with $b > 0$ and a cutoff u. Its binary rule is $d_u = \mathbf{1}\{a + bX + g\zeta > u\}$. Conditional on the sign ζ, this is a threshold rule for X:

$$d_u = \mathbf{1}\{X > h_\zeta\}, \qquad h_+ = \frac{u - a - g}{b}, \quad h_- = \frac{u - a + g}{b}.$$

The two thresholds differ by $h_- - h_+ = 2g/b$. Therefore $\max\{|h_+|, |h_-|\} \ge |g/b|$, regardless of u. Choose a sign attaining this maximum. For that sign, disagreement with $d^* = \mathbf{1}\{X > 0\}$ contains the interval between 0 and h_ζ, restricted to $[-1,1]$. Under $X = \mathrm{sgn}(T)|T|^\nu$, this interval corresponds to one side of $T = 0$ of length $\min\{|h_\zeta|, 1\}^{1/\nu}$. The chosen sign has probability $1/2$, $S = |\eta - \tau| = c|T|$, and the density of T on either side of zero is $\alpha|t|^{\alpha-1}/2$. Consequently, for every cutoff u,

$$\begin{aligned}\mathbb{E}[S\mathbf{1}\{d_u \ne d^*\}] &\ge \frac{1}{2}\frac{c\alpha}{2}\int_0^{\min\{|h_\zeta|,1\}^{1/\nu}} t^\alpha\, dt \\ &\ge \frac{c\alpha}{4(\alpha+1)}\min\{|g/b|^{(\alpha+1)/\nu}, 1\}.\end{aligned}$$

Apply (39) to this route, whose posterior is η and prevalence is τ:

$$V - F_\beta(d_u) = \frac{w\mathbb{E}[S\mathbf{1}\{d_u \ne d^*\}]}{\beta^2\tau + \mathbb{E}d_u}.$$

The denominator is at most $1 + \beta^2\tau$. On the event following (49), every finite auxiliary fit has $b = \widehat{b}_\lambda > 0$ and $g/b > \delta/\sqrt{n}$. Substitution into the last two displays therefore gives, simultaneously for every $\lambda < \infty$ and $u \in \mathbb{R}$, regret at least

$$\frac{wc\alpha}{4(\alpha+1)(1+\beta^2\tau)}\delta^{(\alpha+1)/\nu} n^{-(\alpha+1)/(2\nu)}$$

for all sufficiently large n. The limiting fit at $\lambda = \infty$ has a constant score. Its only threshold rules are the constant predictions 0 and 1; the better of these has regret $V - w\tau/(1+\beta^2\tau) = w\beta^2\tau^2/(1+\beta^2\tau) > 0$.

Representing the fixed-penalty fit by an auxiliary fit. We first check that the empirical minimizer in (41) exists on the event used above. The unpenalized empirical risk, corresponding to $\lambda = 0$, is continuous and strictly convex on the full-rank event and has an attained minimum by assumption. It is therefore coercive: its minimum increase on a unit sphere around that unique minimizer is positive, and convexity gives at least linear growth beyond that sphere. Convex radial symmetry implies $\mathrm{pen}(v) \geq \mathrm{pen}(0)$. Adding the fixed penalty preserves coercivity and strict convexity, so the fixed-loss fit $(\widehat{a}, \widehat{v})$ exists and is unique.

For $v \neq 0$, radial symmetry makes the penalty gradient parallel to v. Convexity and the minimum at zero give

$$\nabla \,\mathrm{pen}(v) = \lambda(v)v, \qquad \lambda(v) = \frac{\langle \nabla \,\mathrm{pen}(v), v\rangle}{\|v\|_2^2} \geq 0, \qquad \nabla \,\mathrm{pen}(0) = 0.$$

At the fixed-loss minimizer, put $\widehat{\lambda} = \lambda(\widehat{v})$ if $\widehat{v} \neq 0$ and $\widehat{\lambda} = 0$ otherwise. Its first-order equations are

$$\mathbb{P}_n \dot{L}_Y(\widehat{a} + \widehat{v}^\top W) = 0, \qquad \mathbb{P}_n W \dot{L}_Y(\widehat{a} + \widehat{v}^\top W) + \widehat{\lambda}\widehat{v} = 0.$$

These are exactly the equations for the auxiliary quadratic-penalty objective at $\lambda = \widehat{\lambda}$. Convexity makes the equations sufficient for a minimum, and strict convexity on the full-rank event makes that minimum unique. Hence

$$(\widehat{a}, \widehat{v}) = (\widehat{a}_{\widehat{\lambda}}, \widehat{v}_{\widehat{\lambda}}).$$

Although the prescribed penalty is fixed, its representing value $\widehat{\lambda}$ can depend on the training sample. The lower-bound event proved above holds for every auxiliary fit and every cutoff on that same sample, so it applies to this particular fit and proves (44). The analogous population first-order equations identify the population fixed-penalty minimizer with a finite auxiliary minimizer, whose informative slope is positive and irrelevant-feature coefficient is zero. Its affine score is therefore decision-aligned as claimed. □

Checking the conditions for familiar losses. The squared, logistic, and exponential base losses are, respectively,

$$L_y(s) = \frac{1}{2}(s-y)^2, \qquad L_y(s) = \log(1+e^s) - ys, \qquad L_y(s) = e^{(1-2y)s}.$$

They are strictly convex and C^2. Their derivative contrasts $L_1'(s) - L_0'(s)$ are -1, -1, and $-e^{-s} - e^s$, so each satisfies (43). We next verify existence and uniform boundedness of their auxiliary minimizers, as required before Theorem 6.

Put $U = (1, X, \zeta)^\top$, choose M_U with $\|U\|_2 \leq M_U$, and let $\kappa_U = \lambda_{\min}(\mathbb{E}UU^\top) > 0$. For any unit coefficient vector $\vartheta \in \mathbb{R}^3$, $(\vartheta^\top U)^2 \leq M_U|\vartheta^\top U|$, and hence $\mathbb{E}|\vartheta^\top U| \geq \kappa_U/M_U$. Write $Y_\pm = 2Y - 1$ and choose $\varepsilon > 0$ with $\eta \in [\varepsilon, 1-\varepsilon]$. Conditioning on U gives

$$\begin{aligned}\mathbb{E}[(-Y_\pm \vartheta^\top U)_+ \mid U] &= (1-\eta)(\vartheta^\top U)_+ + \eta(-\vartheta^\top U)_+ \\ &\geq \varepsilon|\vartheta^\top U|.\end{aligned}$$

A uniform law of large numbers over the compact unit sphere then implies, with probability tending to one,

$$\inf_{\|\vartheta\|_2 = 1} \mathbb{P}_n[(-Y_\pm \vartheta^\top U)_+] \geq \frac{\varepsilon \kappa_U}{2M_U}.$$

Both logistic and exponential losses satisfy $L_Y(s) \geq (-Y_\pm s)_+$. Writing any nonzero augmented coefficient vector as $\theta = \|\theta\|_2 \vartheta$ therefore bounds its empirical risk below by $\|\theta\|_2 \varepsilon \kappa_U/(2M_U)$ on this event. This proves coercivity and existence of a minimizer. At an auxiliary minimizer, comparison with the zero coefficient vector gives empirical risk at most $\mathbb{P}_n L_Y(0)$, because the quadratic penalty is nonnegative and vanishes at zero. For the two losses this upper bound is $\log 2$ and 1, respectively, so the coefficient norms are bounded uniformly over all $\lambda \geq 0$.

For squared loss, the smallest eigenvalue of $\mathbb{P}_n UU^\top$ is at least $\kappa_U/2$ with probability tending to one. On this event the quadratic empirical risk is coercive and attains its minimum. Comparing an auxiliary minimizer with the zero vector gives $\mathbb{P}_n(Y - \theta^\top U)^2 \leq \mathbb{P}_n Y^2 \leq 1$. Since $(\theta^\top U)^2 \leq 2(Y - \theta^\top U)^2 + 2Y^2$, we obtain

$$\frac{\kappa_U}{2}\|\theta\|_2^2 \leq \mathbb{P}_n(\theta^\top U)^2 \leq 4.$$

Thus every auxiliary minimizer satisfies $\|\theta\|_2^2 \le 8/\kappa_U$. Replacing empirical averages by expectations proves the corresponding population statements. These bounds also apply when the slopes are constrained to zero, so they cover the constant-score minimizers used at $t = 0$.

Proof of Theorem 3. We construct a fixed setup whose hard instances satisfy Assumptions 1–3, and additionally verify the stronger routing and fitting conditions used for the matching upper bound. Here $\mathcal{P}$ contains all laws on the fixed observation space satisfying the three stated assumptions. We fix the plan costs strictly below B, so the population benchmark V_n is defined for every $\mathbb{P} \in \mathcal{P}$. In (22), $F_{\beta,\mathbb{P}}$ is the population value of the fitted policy, and both V_n and this value are evaluated under $\mathbb{P}$. The infimum ranges over measurable selections from the full training sample to $\Pi \times \mathbb{R}$, using the prescribed full-data score fits. Randomized selections are also covered by the same lower-bound arguments. Fix $\lambda = 1/4$ and set

$$\ell(y, W; (a, v)) = \tfrac{1}{2}(y - a - v^\top W)^2 + \tfrac{\lambda}{2}\|v\|_2^2,$$

with the clipped-linear link. This single loss serves only to witness sharpness within the permitted class; its penalty is not tuned. Use the common two-router class and three-dimensional augmented design $(1, Q_j, \zeta_j)$ from the proof of Theorem 5, with fixed one-cell partition $\chi(Z) \equiv 1$. Each router selects one fixed action. In the compression construction, set $Q_1 = Q_2 = X$ and $\zeta_1 = \zeta_2 = \zeta$. Thus both routes supply exactly the same numerical training sample. Their strictly convex fitting objectives have the same unique minimizer, so selecting the other router cannot avoid an error shared by this fit.

Fitting regularity. Let $U = (1, Q_j, \zeta_j)^\top$ be the augmented numerical design. Its coordinates are bounded, and its Gram matrix $\mathbb{E}UU^\top$ has a common positive lower eigenvalue in the boundary, route, and compression constructions. For the chosen loss, the population Hessian is

$$\mathbb{E}UU^\top + \lambda\,\mathrm{diag}(0, 1, 1),$$

so the population objective is uniformly strongly convex. The population coefficients computed below have intercept in $(0, 1)$, informative slope at most $1/2$ in absolute value, and sign coefficient zero. They therefore lie in the interior of the same compact convex domain $\Theta_a = [-2, 2]^3$, at a uniform positive distance from its boundary. On this domain, the loss is differentiable and differentiation passes under expectation. Bounded features and coefficients bound the loss gradients uniformly. Their Lipschitz constant is bounded by the operator norm of the per-item Hessian $UU^\top + \lambda\,\mathrm{diag}(0, 1, 1)$, which is uniformly bounded as well. The clipped-linear link is 1-Lipschitz. This verifies all five conditions of Assumption 5 in Appendix A.1, uniformly in n. Lemma 1 gives $a_n = O_p(n^{-1/2})$. The coefficient consistency proved in Lemma 1, together with the positive distance of the population coefficients from the boundary of $[-2, 2]^3$, puts every full-data and fold fit in the interior with probability tending to one. At such an empirical minimizer, the unconstrained gradient is zero. Strict convexity then identifies it with the unique unconstrained fit, so Theorem 6 applies to the full-data fit.

Score conditions for the fixed loss. In the boundary experiment, let $\theta \in \{-1, 1\}$ denote the testing alternative. The population fitted probability is $\bar{p}_\theta(X) = \psi(\rho_\theta + b_\delta X)$, with the clipped-linear link ψ and

$$b_\delta = \frac{1 - \delta^\alpha}{2(1 - \delta^\alpha + \lambda)}.$$

The irrelevant sign has coefficient zero. The coefficients b_δ are uniformly bounded away from zero. Choose $\bar{\tau}_\theta = \rho_\theta - \theta\delta^\nu$. Clipping preserves the oracle decisions at this interior cutoff. At $X = 0$, the score distance $\bar{S}_\theta$ from $\bar{\tau}_\theta$ equals δ^ν, the posterior distance S_θ from the oracle cutoff lies in $[\delta/2, 2\delta]$, and the point has probability δ^α. At $X = \pm 1$, both distances are bounded below uniformly. Thus every nonempty local margin event again consists only of $X = 0$. The calculations for the boundary construction in Appendix B.1 apply with these same masses and distances, verifying decision alignment in Assumption 2 and all three inequalities in Assumption 3.

In the compression experiment, both duplicate routers have the same population score $\bar{p}_\pi = \tau + bX$, with

$$b = \frac{c\,\mathbb{E}|T|^{\nu+1}}{\mathbb{E}|T|^{2\nu} + \lambda} > 0.$$

Clipping is inactive. At cutoff $\bar{\tau}_\pi = \tau$, its decision is $\mathbf{1}\{X > 0\}$ and $\bar{S}_\pi = b|T|^\nu = (b/c^\nu)S_\pi^\nu$. Here $S_\pi = c|T|$, and $b > 0$ is fixed because the loss and this distribution are fixed. The calculation above

with b replacing b_{LS} therefore proves all three inequalities in (16), with constants independent of n. In the route experiment, writing $Q_j = 2X_j - 1$, the population score is

$$\bar{p}_j = \tfrac{1}{2} + \frac{1 - 2e_j}{2(1+\lambda)} Q_j.$$

The positive coefficient of Q_j implies that the score cutoff $\bar{\tau}_j = 1/2$ gives $\mathbf{1}\{Q_j > 0\} = X_j$, the oracle decision. This score cutoff and the posterior cutoff $\tau_j = (1 - e_j)/(1 + \beta^2)$ produce the same predictions, even though their numerical values can differ. Both $|\bar{p}_j - \bar{\tau}_j|$ and $|p_j - \tau_j|$ are bounded below by positive constants for all sufficiently small Δ. Choosing the local radii below these bounds makes the posterior-margin and relative-score-margin events empty. This verifies Assumptions 2 and 3 for the route construction as well.

Routing conditions and the lower bound. Each experiment consists of independent full-information observations, with prevalence bounded away from zero and equal deterministic route costs strictly below B. Collection and deployment have the same law, and each plan has its own probability model. Thus Assumption 1 holds. Set $\mathcal{J}_0 = \mathcal{J}$, corresponding to $\Pi_0 = \Pi$ in Assumption 4. The duplicate routes tie in the boundary and compression experiments, while the route experiment has gap $c_* n^{-\gamma}$. This verifies Assumption 4 and the fixed cost slack in Theorem 2. Finite maxima and minima give common constants across all the instances.

The boundary and route lower bounds in Theorem 5 allow arbitrary learned terminal rules, so they apply in particular to routers and cutoffs selected after fitting this loss. Their exponents are $(1 + \alpha)/(2 + \alpha)$ and, when $\gamma \geq 1/2$, γ. For the compression distribution, the event in (44) bounds regret below by order $n^{-(1+\alpha)/(2\nu)}$ simultaneously over every cutoff of the unconstrained full-data fit. The probability that a coefficient constraint is active tends to zero, so intersecting with the event that the constrained and unconstrained fits coincide preserves the positive lower limit of this probability. Both duplicated routers have that same fit; selecting between them cannot avoid this event.

For $0 < \gamma < 1/2$, consider the two boundary sequences and the constant compression sequence. For $\gamma \geq 1/2$, also include the two route-testing sequences. Every sequence uses the same loss, partition, numerical feature map, and router class just verified. At each sample size n, their component laws P_n all belong to $\mathcal{P}$. For any data-dependent router and cutoff selection and all sufficiently large n, one of these laws yields a regret lower bound at the slowest applicable rate, whose exponent is $r_{\alpha,\nu,\gamma}$ in (21). Taking the minimum of the finitely many positive regret constants and probability bounds gives constants $c, c_0 > 0$ independent of the selection procedure and n. Enlarging the supremum from these constructed laws to all of $\mathcal{P}$ preserves the lower bound. Taking the infimum over data-dependent router and cutoff selections therefore proves (22). □

The hard instances also satisfy Assumption 4 and the fitting regularity in Assumption 5, with $a_n = O_p(n^{-1/2})$. Thus the construction establishes sharpness within the setting of Theorem 2, even though Theorem 3 takes a supremum over the larger class defined by Assumptions 1–3. Theorem 6 separately shows that the compression obstruction persists for every fixed loss satisfying this appendix's smooth convex loss and radial penalty conditions. Attainment of the upper rate by the implemented search also requires negligible empirical optimization error.

The compression lower bound permits arbitrary selection of the cutoff applied to the full-data fit of the fixed loss. It does not cover learned feature maps or partitions, coordinate-specific penalties, arbitrary sample-size-dependent losses, or postprocessors that use the original features again. In particular it cannot, by itself, certify minimax optimality of an implementation with learned representations and partitions.

D Implementation Details

D.1 Coupled search in the experiments

In the experiments, we set $\beta = 1$. ReCal-Route uses budget B, initializes $\phi = 0$, and at each iteration recomputes predictions using the strict cutoff $\phi/(1 + \beta^2)$ before solving the binary assignment problem. The implementation calls `scipy.optimize.linprog` with `method='highs'`, binary bounds, and integer variables. It retains strict improvements in achieved development F_β,

stores the input ϕ with the assignment, and stops after at most 80 iterations, at absolute ratio tolerance 10^{-9}, or upon an unsuccessful solve. A repeated rounded ϕ alone does not imply convergence. The maximum assignment size is $8 \times 66 = 528$ variables for VLM and $8 \times 28 = 224$ for IMDb. No global certificate over the coupled search is asserted. For both main comparisons in Section 5, the out-of-fold (OOF) selected family, penalty, partition, and plans remain fixed after the full-development refit, and a scalar scan maximizes development F_β on the selected-plan full-fit scores. It evaluates the cutoff immediately below the minimum score and every distinct observed score, using strict thresholding without splitting tied scores; exact objective ties choose the largest cutoff. The saved cutoff is frozen before replay. The shared implementation is `vlm_judge_policies/_helpers/recalib_core.py`; recalibration-class and selection wrappers implement the study-specific grids.

Baseline tuning. The five external families (RouterBench-kNN, Avengers-Pro, Unified Router, WISERouter, and a depth-two cascade) are explicitly adapted to the requested F_β objective. Their transformed reward is $(1+\beta^2)Yy_a - \phi y_a$, with prevalence included in the ratio update. Structural tuning uses pooled inner-fold validation confusion counts. kNN uses $k \in \{5, 10, 40\}$; Avengers-Pro uses 60 clusters and four nearest centroids; Unified Router uses logistic $C = 1$; the cascade uses stage penalties $\{0, .001, .01, .1, 1, 10, 100\}$; WISERouter uses $\{8, 12, 16, 20\}$ clusters. VLM uses the full training budget and five WISERouter repeats. IMDb also tunes training-budget fractions $\{.8, .9, 1\}$ and uses one deployment draw per outer seed. IMDb additionally includes a context-free FrugalML–F_β adaptation, using raw score gates and a finite-grid MILP fitted on all rows of training data at the full requested budget, without inner structural cross-validation. Objective matching does not equalize acquisition rules or terminal decision families. Complete configurations, failures, actual costs, and freeze/audit records accompany the canonical reports; failed points and successful held-out budget overruns are retained.

D.2 Quantile partitions

For each outer development cohort, we construct four candidate directions: one logistic regression of Y on the routing context and three reduced-rank directions for the judge-correctness responses. Each direction is binned separately at its development $1/K, \ldots, (K-1)/K$ quantiles, with nominal $K = 8$. Directions and cutpoints are then frozen for held-out deployment. They are learned on all rows of training data, not on an independent pilot or within the recalibration folds.

For a scalar context score ξ, let $\widehat{b}_1, \ldots, \widehat{b}_{K-1}$ be its development quantiles. The cell is

$$k_i = 1 + \sum_{q=1}^{K-1} \mathbf{1}\{\xi(Z_i) > \widehat{b}_q\}.$$

This agrees with left insertion at tied cutpoints. Repeated scores can leave empty cells; only occupied development cells enter the optimizer. Quantile bins avoid crossing several context scores into a sparse multidimensional grid. The candidate directions are specified next in Appendix D.4.

D.3 Plan library, prediction states, and realized costs

Here j, k index individual judges, and J is the number of judges. For an action a, C_a is its random realized cost and c_{ia} is its recorded value for item i, as in Assumption 1.

Example 1 (Item-dependent cost of a confirmation plan)**.** The plan $a = A \wedge B$ is an ordered call plan: call judge A first, then call judge B only if $y_A = 1$. Suppose context acquisition is free and the fixed call costs are $c_A = 1$ and $c_B = 3$ units. For an item on which A returns 0, the plan stops after A and costs 1 unit. For an item on which A returns 1, it calls B for confirmation and costs 4 units. Thus

$$C_{A\wedge B} = c_A + \mathbf{1}\{y_A = 1\}c_B = 1 + 3\mathbf{1}\{y_A = 1\}.$$

Although the plan and individual call prices are fixed, its realized cost is random because y_A varies across items. Accordingly, $C_{\pi(Z)}$ is the realized cost of executing the selected plan, and $\mathbb{E}[C_{\pi(Z)}]$ is its population mean cost.

Table 3: Executable plan families; y_j is judge j's hard output.

Plan	Second call	Raw decision
Singleton j	Never	y_j
Recovery $j \vee k$	If $y_j = 0$	$y_j \vee y_k$
Confirmation $j \wedge k$	If $y_j = 1$	$y_j \wedge y_k$

On row i, let y_{ij} and q_{ij} be judge j's binary output and confidence; q_{ij} need not be a positive-class probability. A singleton plan reveals state $\{j\}$. A recovery plan $j \vee k$ reveals $\{j\}$ when $y_{ij} = 1$ and $\{j, k\}$ otherwise; confirmation $j \wedge k$ reveals $\{j\}$ when $y_{ij} = 0$ and $\{j, k\}$ otherwise. Call order still determines execution and cost. For risk estimation, pair-state coordinates are ordered by judge index. If $\widetilde{z}_i$ is the context representation fitted on the applicable training rows (four recalibration folds or the full development cohort), the feature rows, including their categorical state, are

$$\begin{aligned} W_{i,\{j\}} &= \big(\{j\}, \widetilde{z}_i, y_{ij}, q_{ij}\big), \\ W_{i,\{r,t\}} &= \big(\{r,t\}, \widetilde{z}_i, y_{ir}, q_{ir}, y_{it}, q_{it}\big), \qquad r < t. \end{aligned} \tag{50}$$

The categorical state selects a separate regression; only the numerical predictors are standardized and included in its additive design, together with an intercept. Thus the $2J^2 - J$ ordered executable plans use only $J + \binom{J}{2}$ state models. In the notation of (10), θ_a collects the coefficients of the state models used by plan a, and Φ_a selects the numerical block for the returned state, setting the other blocks to zero. Plans using the same state model share its fitted coefficients. The pooling restriction posits that, conditional on the state identity, Z, and the called judges' recorded binary outputs and confidence scores, label risk does not depend further on call order. If that pooling restriction is doubtful, call-plan identity should be included as an additional feature.

The six VLM judges use the archived identifiers M1, M2, M3, M4, M5, and V235. Exact checkpoint names require the original collection provenance; adjacent historical records disagree on M2. With J base judges, the implementation uses J singletons and, for every ordered pair $j \neq k$, a recovery and a confirmation plan, for a total of $J + 2J(J-1) = 2J^2 - J$ plans. If c_j is judge j's cost and c_{ctx} is the context cost, then

$$\begin{aligned} C_j &= c_{\mathrm{ctx}} + c_j, \\ C_{j\vee k} &= c_{\mathrm{ctx}} + c_j + \mathbf{1}\{y_j = 0\} c_k, \\ C_{j\wedge k} &= c_{\mathrm{ctx}} + c_j + \mathbf{1}\{y_j = 1\} c_k. \end{aligned}$$

The benchmark ledger uses nominal cost units per 1,000 items: M1 0.100, M2 0.100, M3 1.431, M4 5.675, M5 5.901, and V235 0.100. In the tag-context panels, $c_{\mathrm{ctx}} = 0.100$ accounts for tag acquisition; in the CLIP-context panels, the experimental representation charge is zero. The budget B uses these same units, which we do not interpret as certified currency charges.

D.4 Candidate partition directions

Let z_i be row i's numerical routing-context vector. On row i of training data, let $R_{ij} = \mathbf{1}\{y_{ij} = Y_i\}$ indicate whether judge j is correct. Let Z_c and R_c be the centered routing-context and correctness matrices and set $\widehat{\Omega} = \mathrm{lstsq}(Z_c, R_c)$. If

$$Z_c \widehat{\Omega} = U_R \Sigma_R V_R^\top,$$

write v_{Rh} for the hth column of V_R. Then direction $\widehat{\omega}_h = \widehat{\Omega} v_{Rh}$ gives scalar score $\xi_{ih}^{\mathrm{perf}} = z_i^\top \widehat{\omega}_h$. The risk direction is the coefficient vector of an L2-logistic regression of Y on the routing context, with inverse penalty $C = 1$ and at most 300 iterations; its intercept does not affect quantile ordering. The fixed candidate order is risk first, then three performance directions (nonzero directions where required). Correctness is a response for these routing directions; the recalibration regressions instead target Y.

For VLM, routing uses the six raw tag coordinates or the first 16 components of full-development, centered, unwhitened CLIP PCA; recalibration fitting uses 128 PCA coordinates. Each recalibration fold fits its own PCA, while ridge and logistic reuse exactly matched fold memberships and PCA transforms. IMDb routing uses all 8 tag, 64 LSA, or 72 combined coordinates, without an extra

routing PCA. Directions and cutpoints use development data only, but are label-dependent; the theory must account for that learning rather than treat the four observed partitions as an independently fixed class.

D.5 Five-fold regularized recalibration

Let T be the rows used for the current recalibration fit: the training folds for a fold fit, or all development rows for the full-data fit. For state e, let $x_{i,e}$ be the numerical predictor design formed from (50), excluding the intercept and state identity, standardized using statistics computed on T; near-constant columns are assigned scale one. The two VLM objectives are

$$(\widehat{a}_e, \widehat{b}_e)_{\text{ridge}} \in \arg\min_{a,b} \left\{ \frac{1}{2} \sum_{i \in T} (Y_i - a - x_{i,e}^\top b)^2 + \frac{\lambda_T}{2} \|b\|_2^2 \right\},$$

$$(\widehat{a}_e, \widehat{b}_e)_{\text{logistic}} \in \arg\min_{a,b} \left\{ \sum_{i \in T} \left[\log(1 + e^{a + x_{i,e}^\top b}) - Y_i (a + x_{i,e}^\top b) \right] + \frac{\lambda_T}{2} \|b\|_2^2 \right\}.$$

Here $\widehat{a}_e$ is the fitted intercept and $\widehat{b}_e$ is the fitted slope vector. For $e = \{j\}$ or $e = \{j, k\}$ with $j < k$, the resulting function $\psi(\widehat{a}_e + x_e^\top \widehat{b}_e)$ is the model $\widehat{p}_j$ or $\widehat{p}_{jk}$, respectively, from Section 5; x_e is the new item's numerical feature vector transformed using the statistics fitted on T. The coefficient vector $\widehat{\theta}_a$ in Section 4 collects the fitted blocks used by plan a. Both objectives leave the intercept unpenalized and use no class weighting. Ridge predictions are clipped to $[0, 1]$; logistic predictions use the sigmoid. A logistic link alone does not establish probability calibration. All state models are trained on the complete prediction panel for their training rows, not only on rows where a particular gate continues.

VLM. The additive design in Equation (50) has no engineered interactions. Tags give 8/10 nonintercept singleton/pair predictors; CLIP gives 130/132. The common policy-level grid is $\Lambda = \{10^{-1+6r/7} : r = 0, \ldots, 7\}$. One penalty is shared across all 21 states of a candidate. Fold fits use $\lambda_T = \lambda |T|/n$; the full-development fit uses λ. Equal numerical penalties across the two losses do not imply equal effective shrinkage. Outer seed s uses five shuffled stratified recalibration folds with seed $20260921 + 1000s + 999$. Each fold fits PCA, state scaling, and regression on its training rows. Linear ridge uses centered SVD; logistic uses the Newton–Cholesky solver, inverse penalty $C = 1/\lambda_T$, tolerance 10^{-9}, and at most 100 iterations. Nonconvergence is an error. At every budget, each family compares eight penalties and four partitions; ReCal-Route compares their union. Exact ties prefer ridge, then increasing penalty, then partition order.

IMDb. The four services give ten prediction states and 28 plans. The declared class has 16 ridge and seven L2-logistic candidates. Ridge uses the same eight numerical penalties for each of two designs: additive raw predictions and prediction interactions. Logistic uses additive designs with $C \in \{0.01, 1, 100\}$, interaction designs with $C \in \{0.01, 1\}$, a verdict-only design with $C = 1$, and a no-context interaction design with $C = 1$. For each acquired judge, interactions add $y_j q_j$ and q_j^2; pair designs additionally add $y_1 y_2, q_1 q_2, y_1 q_2, y_2 q_1$. There are no interactions with context. The no-context candidate excludes context from recalibration fitting only; routing still uses it. Thus the two IMDb candidate classes differ in design as well as loss. Candidate-order ties retain raw ridge, interaction ridge, then logistic in the order just given.

IMDb uses a training-row `StandardScaler`, Cholesky ridge, and logistic L-BFGS with tolerance 10^{-6} and at most 1,000 iterations. Unlike VLM, the same numerical ridge parameter `alpha` or logistic C is used in each fold and the full-development fit, without sample-size rescaling. Inner shuffled stratified group folds use the outer seed and keep normalized duplicate reviews together. Every review is scored once by a fit excluding its group. Text TF–IDF/SVD is fitted separately in each recalibration-training fold; the full-development representation is used for routing partitions and deployment, not for producing OOF predictions.

The text representation uses word unigrams/bigrams, minimum document frequency 3, maximum fraction 0.98, at most 30,000 terms, sublinear term frequency, and L2-normalized TF–IDF. A randomized 64-component SVD (seven iterations, seed 1729) is row-normalized. Combined context concatenates these 64 coordinates and eight fixed text-rule tags. Raw service scores remain numerical

features with no assumed probability semantics. Costs per 10,000 reviews are Google 2.5, Baidu 3.5, Amazon 0.75, and VADER 0.001; local text representation has no acquisition charge. These historical units differ from VLM's ledger.

Full-development deployment. Each recorded deployment uses a fit on all rows of training data; fold models are not averaged. The two main comparisons apply the final full-fit cutoff scan with the selected router fixed.

The recorded procedure saves budget-specific deployment configurations, including the selected recalibration family, feature specification, and penalty. It contains the context preprocessing parameters (PCA for VLM, TF–IDF/SVD for IMDb), selected partition score and bin edges, cell–plan map, all $J + \binom{J}{2}$ prediction-state feature transformations and coefficient vectors, decision cutoff, and budget-specific solver diagnostics. If a bin contains no development observations but can receive a future context, its saved fallback plan is the first declared singleton (M1 in VLM). A plan that stops after one call uses the corresponding singleton risk model; a plan that continues uses the corresponding pair-state model.

Table 4: Settings of the recorded cross-fitted replay experiments.

Component	VLM	IMDb
Outer protocol	100 stratified 70/30 splits	40 grouped stratified 80/20 splits
States / plans	21 / 66	10 / 28
Routing class	$K = 8$; one logistic + three RRR directions	Same nominal class
Recalibration folds	Five stratified; fold-local PCA/scaling	Five grouped; fold-local text/scaling
Recalibration candidates	Eight penalties per family; additive	16 ridge + seven logistic
Context dimensions	Tags 6; CLIP routing 16 / recalibration 128	Tags 8 / LSA 64 / combined 72
Selection	Per-budget development F_1	Same criterion
Coupled search	At most 80 iterations; tolerance 10^{-9}	Same routine
Budgets	Ten points on $[0, 2]$; zero infeasible	14 points from 0.001 to 6

E Additional Empirical Results

E.1 Evaluation protocol

Data and evaluation. IMDb contains 25,000 reviews and four archived sentiment services from the public FrugalML panel, with positive prevalence $1/2$. Its 40 grouped 80/20 splits keep duplicate texts together in the outer split and recalibration folds. Context is eight text tags, 64-dimensional TF–IDF/LSA features, or their concatenation. The VLM study uses 3,788 AI-generated 3D items and six judges, with six content tags or frozen OpenCLIP features as context. We evaluate 100 stratified 70/30 development/test splits. We optimize and evaluate F_1 in both studies.

Methods. ReCal-Route selects ridge or logistic recalibration models using five-fold cross-fitted development F_1, with representations and scalers fitted within each recalibration fold. After full-development refitting, the final cutoff is optimized with the model and router fixed (Algorithm 1). The raw variant optimizes routing with raw Boolean decisions. Baselines are F_1 adaptations of RouterBench-kNN, Avengers-Pro, Unified Router, WISERouter, and a depth-two cascade; IMDb also includes FrugalML. Candidate designs and tuning grids are in Appendix D.

Comparison scope. Within each context, the first four baselines share our singleton and gated two-service plans: 28 for IMDb and 66 for VLM. Cascade and FrugalML retain their own two-call mechanisms. All methods charge actual calls. IMDb has no representation surcharge, and FrugalML is context-free. VLM representation charges are matched across methods: 0.1 for tags and zero for CLIP. Requested budgets use historical nominal units per 10,000 reviews or 1,000 VLM items. Methods retain different policy classes and cost estimators, so equal caps need not give equal realized expenditure.

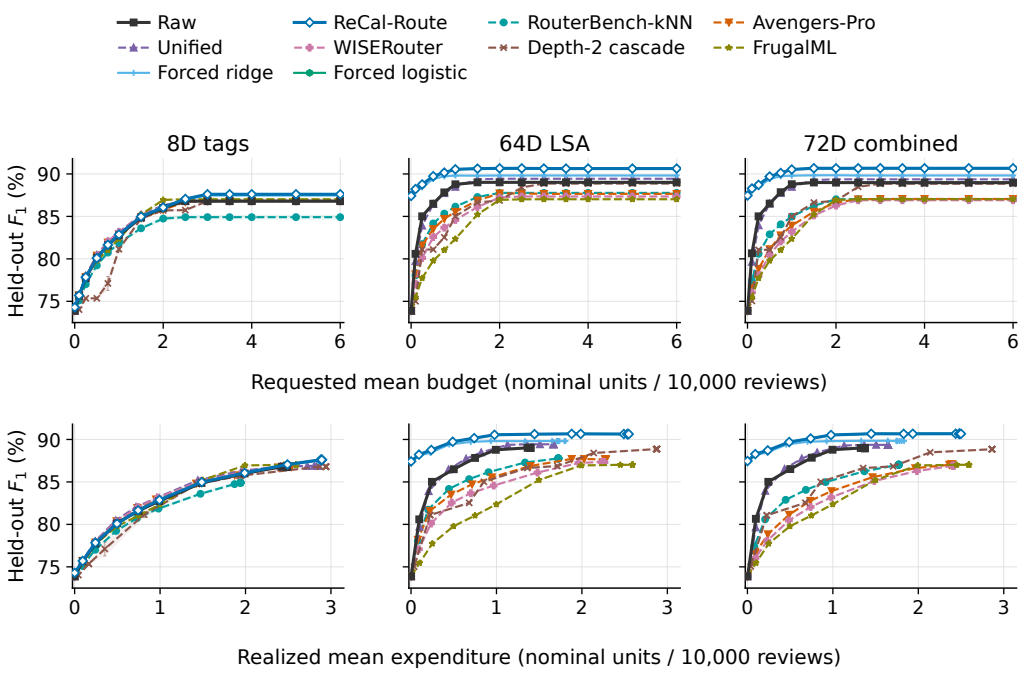


Figure 2: IMDb F_1 across tags, LSA, and combined context: requested budget (top) and realized mean expenditure (bottom), over 40 grouped splits. Intervals are descriptive pointwise Student-t 95% intervals. Lower panels connect measured points in requested-budget order. Score intervals placed at mean expenditure do not condition on a fixed realized cost.

E.2 Complete budget curves

The figures below show the complete requested-budget comparisons for F_1, with separate VLM tag and embedding contexts. IMDb also shows realized expenditure and the forced ridge and logistic candidate classes. All curves connect recorded evaluations; no extension tails are added.

ReCal-Route improves mean F_1 over raw routing at all 54 positive-budget/context/input combinations: 0.02–3.59 pp with tags and 2.06–5.07 pp with CLIP (Figures 3 and 4). It also has the highest available-case mean F_1 at every positive budget in both render settings. For multimodal inputs, RouterBench-kNN has higher mean F_1 than ReCal-Route with CLIP throughout, whereas their ordering reverses with tags. These contexts differ in both features and representation charges (0.1 for tags; zero for CLIP, matched across methods).

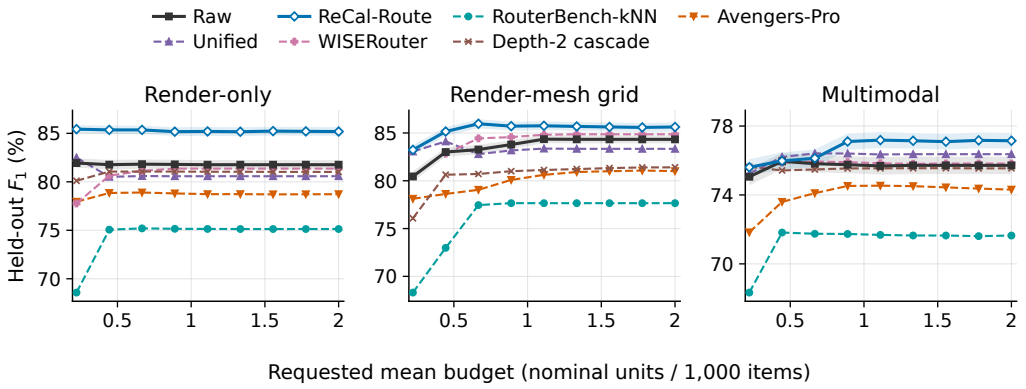


Figure 3: VLM F_1 with 6 content tags across the three judge-input settings. Solid curves are raw and ReCal-Route; dashed curves are context-matched benchmarks. The axis gives requested budget, including the common tag charge. Means use available splits out of 100; shading is descriptive mean $\pm 1.96\,\mathrm{SD}/\sqrt{n_{\mathrm{valid}}}$ for raw routing and ReCal-Route. Infeasible zero-budget points are omitted; availability is described below.

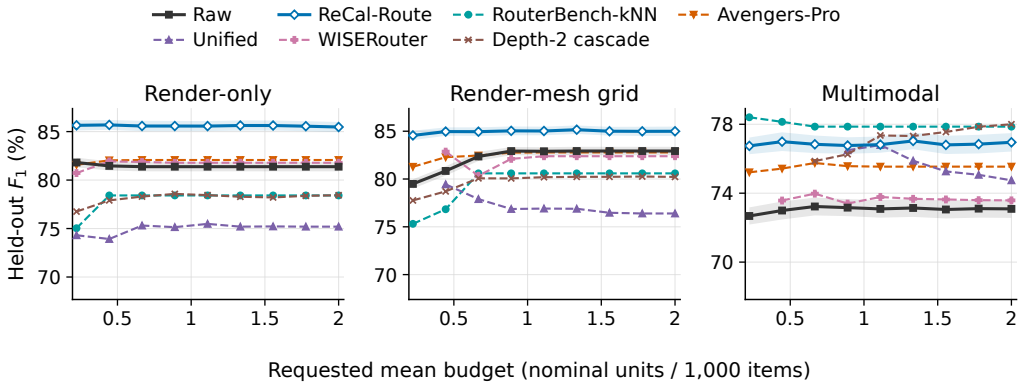


Figure 4: VLM F_1 with CLIP context and zero representation surcharge for every method. Layout, intervals, and corresponding score axes match Figure 3. ReCal-Route improves F_1 over raw routing throughout, while benchmark rankings depend on the input setting. Means can use different successful split subsets.

E.3 Endpoint comparisons

Tables 1, 2, and 5 give the largest-budget operating points for VLM with tags, VLM with CLIP embeddings, and IMDb, respectively. All use the final full-fit cutoff for recalibrated methods. In VLM, ReCal-Route exceeds raw routing in all six context/input settings. In IMDb, ReCal-Route and logistic select the same endpoint policies. The LSA logistic–ridge gap is 0.817 pp.

Table 5: IMDb mean held-out F_1 (%) at $B = 6$ per 10,000 reviews, with the final full-fit cutoff for recalibrated methods. All entries use 40 matched splits. ReCal-Route and logistic select the same policies at this endpoint.

Method	Tags	LSA	Combined
Raw	86.792	89.011	88.979
Ridge	87.464	89.815	89.819
Logistic	87.606	90.633	90.667
ReCal-Route	87.606	90.633	90.667

E.4 Availability and reporting conventions

VLM uses ten equally spaced budgets on $[0, 2]$; zero-budget outcomes are infeasible and omitted from the plots. Across the positive-budget grid, external-method coverage ranges from 43.41% to 99.89% by method and context, versus 100% for raw routing and ReCal-Route. WISERouter contributes a seed/budget outcome only when all five of its repetitions succeed; its seed-level score is their mean. Each displayed point then averages the available seed-level scores. The two VLM contexts have separate benchmark fits, and external tag adaptations use the same six binary tags plus a constant coordinate. All tag methods incur the same 0.1 representation charge; embedding methods incur none. Corresponding tag and embedding panels use the same score-axis limits.

IMDb uses 14 budgets from 0.001 to 6 per 10,000 reviews. WISERouter has 161 unsuccessful outcomes out of 1,680; the other external methods complete every point. The four external benchmarks sharing our 28-plan library have 424 successful held-out budget overruns, all retained. FrugalML is context-free, so its repeated curve in each context panel represents the same experiment. Budget-grid comparisons with Unified Router average all fourteen budgets equally within each split and then average the forty paired split means. These are summaries over the declared budget grid, not areas under an interpolated frontier. The score gains do not adjust for the accompanying differences in realized expenditure.

The figures average held-out scores within each method's available successful split subset, rather than pooling confusion counts across splits. Different availability can change the contributing subset; ordering mean curves is not a statement about paired win proportions. VLM bands are mean $\pm 1.96\,\mathrm{SD}/\sqrt{n_{\mathrm{valid}}}$; IMDb uses pointwise Student-t 95% intervals. Both describe variability across overlapping reseeds of previously examined datasets. Successful cost overruns remain in the means, and unavailable outcomes are not replaced with outcomes from another method or budget.

E.5 Additional selection diagnostics

At the largest budget, logistic exceeds ridge in all three IMDb contexts (Table 5). On IMDb, ReCal-Route chooses logistic at 1,495 of 1,680 operating points and ridge at 185. The final full-fit cutoff scan preserves these family choices. The candidate classes comprise 16 ridge and seven regularized logistic models with the designs specified in Appendix D.5; their comparison changes both design and loss.