Uncertainty Quantification of Click and Conversion Estimates for the Autobidding

Ivan Zhigalskii[3] **Andrey Pudovikov**[1,2] **Aleksandr Katrutsa**[1] **Egor Samosvat**[4]

[1]AI Center, Lomonosov Moscow State University, Moscow, Russia,
[2]IAI, Lomonosov Moscow State University, Moscow, Russia,
[3]HSE, Moscow, Russia,
[4]Independent researcher, Moscow, Russia,

Abstract

Modern e-commerce platforms employ various auction mechanisms to allocate paid slots for a given item. To scale this approach to millions of auctions, the platforms suggest promotion tools based on the autobidding algorithms. These algorithms typically rely on the Click-Through-Rate (CTR) and Conversion-Rate (CVR) estimates from a pre-trained machine learning model. However, the predictions of such models are uncertain and can significantly affect the performance of the autobidding algorithm. To address this issue, we propose the `DenoiseBid` method, which corrects the generated CTRs and CVRs to make the resulting bids more efficient in auctions. The underlying idea of our method is to employ a Bayesian approach and replace noisy CTR or CVR estimates with those from recovered distributions. To demonstrate the performance of the proposed approach, we perform extensive experiments on the synthetic, iPinYou, and BAT datasets. To evaluate the robustness of our approach to the noise scale, we use synthetic noise and noise estimated from the predictions of the pre-trained machine learning model.

1 INTRODUCTION

Modern e-commerce platforms allocate advertising slots through automated auction mechanisms that process millions of auctions daily. To operate at this scale, platforms deploy autobidding systems that compute optimal bids for advertisers. These systems solve constrained optimization problems parameterized by Click-Through Rates (CTR) and Conversion Rates (CVR), which are estimated from data.

Under the widely studied Second-Price Auction (SPA) rule Yang et al. [2019], Aggarwal et al. [2024], He et al. [2021], Zhang et al. [2017], Khirianova et al. [2025], Pudovikov et al. [2025, 2026], the optimization problem for autobidding systems can be stated as a linear programming (LP) problem. From such a problem follows that the optimal bid is a linear function of CTR and CVR Yang et al. [2019], Aggarwal et al. [2024]. Consequently, estimation errors in these quantities propagate directly into the bid, potentially leading to suboptimal budget allocation and violations of the advertiser's cost-per-click (CPC) constraints. We focus on SPA because the resulting linear dependence of bid on external quantities, e.g., CTR and CVR, allows Bayesian treatment of uncertainties. The extension to first-price auctions is an important direction that requires alternative approaches and is left for future work.

In practice, CTR and CVR are predicted by machine learning models whose outputs are inherently uncertain. This motivates the development of a bidding framework that explicitly accounts for predictive noise. We propose `DenoiseBid`, a Bayesian autobidding method that replaces noisy estimates of CTR and CVR with their posterior expectations, conditioned on the model's predictions and a prior distribution recovered from historical data.

The main contributions of this study are as follows:

1. We state the autobidding problem under noisy CTR and CVR values and derive a closed-form bidding rule based on their posterior expectations.
2. We develop the `DenoiseBid` method, which recovers the prior distribution of CTR and CVR from observations and computes denoised bids in closed form.
3. We perform extensive empirical validation of `DenoiseBid` on four datasets under both synthetic and empirical noise and demonstrate consistent gains.

2 RELATED WORKS.

Autobidding formulations. The autobidding problem can be viewed as a specialization of Real-Time Bidding (RTB) Ou et al. [2023] in which the platform optimizes

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:8253–8269.

bids under global performance goals. Existing approaches employ optimization Aggarwal et al. [2024], Zhang et al. [2017], reinforcement learning Cai et al. [2017], He et al. [2021], or control techniques Yang et al. [2019], Stram et al. [2024]. To the best of our knowledge, Yang et al. [2019] were the first to cast the advertiser's conversion-maximization problem with joint budget and CPC constraints as an LP; the formulation and its dual-based bidding rule have since become a standard baseline Aggarwal et al. [2024], Pudovikov et al. [2026].

Uncertainty quantification for CTR/CVR prediction. The LP-based bids depend on estimated CTR and CVR, and the autobidding problem can be formulated through a Bayesian perspective Aggarwal et al. [2024]. From the modeling side, CTR/CVR prediction spans tabular feature-interaction models and deep behavior-sequence architectures Zhu et al. [2022]. For gradient-boosted decision trees (GBDT), virtual ensembles decompose predictive variability into data and knowledge components Malinin et al. [2021]. Similar ideas for deep models include deep ensembles and variational Bayesian methods Abdar et al. [2021]. Hence, calibrated uncertainty estimates are available for the dominant model families used in practice.

Propagating uncertainty into bidding. The critical remaining question is how to incorporate these calibrated uncertainties into the autobidding layer. Shih et al. [2023] cluster CTR values and combine them with reinforcement learning to construct a bidding policy; however, their clustering-based discretization can lose information when the CTR distribution is unimodal or highly concentrated. Zhang et al. [2017] introduce a risk-management framework that models CTR uncertainty via Bayesian logistic regression; while principled, this restricts the prediction model to a specific parametric family, precluding the use of modern GBDT or deep-learning predictors Zhu et al. [2022]. The recent `RobustBid` Pudovikov et al. [2026] applies robust optimization with uncertainty sets constructed from quantile estimates of CTR and CVR distributions, achieving strong constraint satisfaction; however, it does not incorporate per-prediction uncertainty estimates from the ML model.

`DenoiseBid` jointly leverages the empirical distribution of CTR and CVR values and per-prediction uncertainty estimates to construct denoised bids (see Table 1). Unlike prior methods, it is model-free, retains a closed-form bidding rule, and exploits Bayesian posterior expectations.

Table 1: Comparison `DenoiseBid` with other methods.

Reference	Multiple uncertainty sources	Usage of CTR distribution	UQ of prediction model
[Shih et al., 2023]	✗	✓	✗
[Zhang et al., 2017]	✗	✓	✗
[Pudovikov et al., 2026]	✓	✓	✗
This paper	✓	✓	✓

3 PROBLEM STATEMENT

This section formalizes the deterministic autobidding optimization problem under perfect knowledge of click and conversion probabilities. The resulting bidding formula serves as the baseline that Section 4 extends to the Bayesian setting.

Consider a single advertiser participating in a sequence of T single-slot second-price auctions (SPAs). For each auction $t \in [T] = \{1, \dots, T\}$, we denote by CTR_t the click-through rate, by CVR_t the conversion rate, by $wp_t > 0$ the realized winning price, and by $bid_t \geq 0$ the advertiser's bid. We adopt an offline setting in which the winning prices $\{wp_t\}_{t=1}^T$ and the ground-truth probabilities $\{CTR_t, CVR_t\}_{t=1}^T$ are known. Under SPA rules, the advertiser wins auction t whenever $bid_t \geq wp_t$ and pays wp_t.

The advertiser seeks to maximize the total expected number of conversions subject to a budget constraint B and a target cost-per-click (CPC) limit C, both given in advance. Following Yang et al. [2019], Aggarwal et al. [2024], we relax the binary allocation variables $x_t \in \{0, 1\}$ to $x_t \in [0, 1]$, yielding the following LP problem:

$$\begin{aligned} &\max_{0 \leq x_t \leq 1} \quad \sum_{t=1}^{T} x_t \cdot CTR_t \cdot CVR_t \\ &\text{s.t.} \sum_{t=1}^{T} x_t \cdot wp_t \leq B, \\ &\quad \frac{\sum_{t=1}^{T} x_t \cdot wp_t}{\sum_{t=1}^{T} x_t \cdot CTR_t} \leq C. \end{aligned} \tag{1}$$

Note that it is possible to extend the CPC constraint to a more general case with Return On Investment constraints He et al. [2021]; however, the CPC formulation is more widespread in the literature Aggarwal et al. [2024].

The optimal bid is derived from the Lagrangian dual problem of (1) with dual variables $p \geq 0, q \geq 0$ corresponding to budget and CPC constraints. Complementary slackness conditions help to derive the closed-form expression

$$bid_t = \frac{1}{p^* + q^*} CVR_t \cdot CTR_t + \frac{q^*}{p^* + q^*} C \cdot CTR_t, \tag{2}$$

where p^* and q^* are the optimal dual variables. The detailed derivation follows the same scheme as in Yang et al. [2019], Aggarwal et al. [2024] and is extended to stochastic modification of problem (1) in Section 4.

Equation (2) shows that bid_t is a linear function of CTR_t and CVR_t. Therefore, estimation errors in these quantities propagate directly into the bid and can affect the bidding strategy's performance. Section 4 extends the optimization problem (1) to handle such uncertainty from the Bayesian perspective and improve bidding performance.

4 DENOISEBID METHOD

We assume the bidding agent observes noisy estimates of the ground-truth click and conversion probabilities. However, the agent's objective is to maximize the number of actual conversions. Therefore, the core idea of DenoiseBid is to shift from deterministic to stochastic optimization, maximizing the conditional expectation of the true conversion probability given the noisy observations. Formalization of this idea leads to the following optimization problem:

$$\begin{aligned} \max_{0\leq x_t\leq 1} \; & \mathbb{E}\left[\sum_{t=1}^{T} x_t \cdot CTR_t \cdot CVR_t \middle| \mathcal{O}\right], \\ \text{s.t.} \quad & \sum_{t=1}^{T} x_t \cdot wp_t \leq B, \\ & \mathbb{E}\left[\frac{\sum_{t=1}^{T} x_t \cdot wp_t}{\sum_{t=1}^{T} x_t \cdot CTR_t} \middle| \mathcal{O}\right] \leq C, \end{aligned} \tag{3}$$

where the expectation is taken over the latent random variables $\{CTR_t, CVR_t\}_{t=1}^{T}$, conditioned by observations $\mathcal{O} = \left\{\left(\widehat{CTR}_t, \widehat{CVR}_t\right)\right\}_{t=1}^{T}$.

To reduce the stochastic optimization problem to LP, we employ a first-order Taylor approximation for the expectation of a ratio in the average cost-per-click constraint:

$$\frac{\sum_{t=1}^{T} x_t \cdot wp_t}{\mathbb{E}\left[\sum_{t=1}^{T} x_t \cdot CTR_t \middle| \mathcal{O}\right]} \leq C. \tag{4}$$

By rearranging the terms, we arrive at the canonical primal formulation:

$$\begin{aligned} \max_{0\leq x_t\leq 1} \quad & \mathbb{E}\left[\sum_{t=1}^{T} x_t \cdot CTR_t \cdot CVR_t \middle| \mathcal{O}\right], \\ \text{s.t.} \quad & \sum_{t=1}^{T} x_t \cdot wp_t \leq B, \\ & \sum_{t=1}^{T} x_t \cdot \left(wp_t - C \cdot \mathbb{E}\left[CTR_t | \mathcal{O}\right]\right) \leq 0. \end{aligned} \tag{5}$$

This representation demonstrates that the agent optimizes the posterior mean of conversions while accounting for the target CPC based on the posterior mean of the CTR.

Similarly to Section 3, we use the dual problem to (5) and complementary slackness conditions to obtain the optimal bidding formula:

$$\begin{aligned} bid_t = & \frac{1}{p^* + q^*}\mathbb{E}[CTR_t\, CVR_t \mid \mathcal{O}] + \\ & + \frac{q^*}{p^* + q^*} C \cdot \mathbb{E}[CTR_t \mid \mathcal{O}], \end{aligned} \tag{6}$$

where p^* and q^* are the optimal dual variables associated with the budget and CPC constraints, respectively. The complete derivation of (6) is provided in Appendix A.

Consequently, the problem of optimal bidding reduces to computing conditional expectations. Applying Bayes' formula, we write out the expectation of a latent variable X (representing CTR or $CTR \cdot CVR$) in terms of the likelihood and the prior distribution:

$$\begin{aligned} \mathbb{E}[X \mid \mathcal{O}] = & \int_X X \cdot p(X \mid \mathcal{O}) \mathrm{d}X = \\ & = \frac{\int_X X \cdot p(\mathcal{O} \mid X) p(X) \mathrm{d}X}{\int_X p(\mathcal{O} \mid X) p(X) \mathrm{d}X}. \end{aligned} \tag{7}$$

Thus, calculating the posterior expectation requires the following components: a noise model (likelihood) $p(\mathcal{O} \mid X)$ and a prior distribution $p(X)$, which captures the distribution of the true values. In the subsequent sections, we analyze the calculation of these expectations under different specifications, addressing CTR-only and joint CTR-CVR uncertainty scenarios. As a result, we derive expressions for the experimental evaluation of the DenoiseBid method.

4.1 BIDDING UNDER CTR-ONLY UNCERTAINTY

To illustrate the mechanism of DenoiseBid, we first consider a simple case where noise is concentrated in CTR predictions, i.e., $\widehat{CVR}_t = CVR_t$. Furthermore, we assume that CTR and CVR predictions are independent of each other and across different auctions. In this case, the joint posterior expectation in (6) decomposes as:

$$\begin{aligned} \mathbb{E}\left[CTR_t \cdot CVR_t \middle| \left\{\left(\widehat{CTR}_t, \widehat{CVR}_t\right)\right\}_{t=1}^{T}\right] = & \\ = CVR_t \cdot \mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right]. & \end{aligned} \tag{8}$$

Consequently, the optimal bid from (6) is directly proportional to this posterior expectation:

$$bid_t = \frac{CVR_t + q^* C}{p^* + q^*} \mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right]. \tag{9}$$

In turn, the expectation is equal:

$$\begin{aligned} & \mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] = \\ & = \frac{\int CTR_t \cdot p\left(\widehat{CTR}_t | CTR_t\right) p(CTR_t) \mathrm{d}CTR_t}{\int p\left(\widehat{CTR}_t | CTR_t\right) p(CTR_t) \mathrm{d}CTR_t}. \end{aligned} \tag{10}$$

Noise model. While we do not assume that estimation errors are strictly Gaussian in all practical scenarios, modeling noise as additive and normally distributed in the logit space is a well-motivated and tractable approximation Bishop and Nasrabadi [2006]. This choice is particularly natural for CTR predictors, which typically generate a real-valued score before applying a sigmoid transformation. Furthermore, characterizing the logit distribution in terms of its first two moments provides a standard basis for quantifying predictive uncertainty Kendall and Gal [2017].

Formally, since click probabilities $CTR_t \in [0,1]$, we perform the estimation in the logit space by defining the logit-CTR ξ_t and its noisy observation $\hat{\xi}_t$:

$$\xi_t = \ln \frac{CTR_t}{1 - CTR_t}, \quad \hat{\xi}_t = \ln \frac{\widehat{CTR}_t}{1 - \widehat{CTR}_t}. \tag{11}$$

We assume the observation model as:

$$\hat{\xi}_t = \xi_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{12}$$

where $\{\varepsilon_t\}_{t=1}^T$ are independent random variables and independent on the ground-truth logit-CTRs ξ_t. Then the likelihood can be written as:

$$p\left(\hat{\xi}_t \,\middle|\, \xi_t\right) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(\hat{\xi}_t - \xi_t)^2}{2\sigma_t^2}}. \tag{13}$$

Gaussian mixture prior. Real-world CTR distributions are often complex and multimodal. Therefore, we model the prior distribution of the logit-CTR ξ as a Gaussian Mixture:

$$\xi \sim \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(\mu_k, \theta_k^2\right), \tag{14}$$

where K is number of mixture components, π_k, μ_k and θ_k^2 are the weight, mean and variance of the k-th component, respectively. `DenoiseBid` with this prior adapts to the empirical distribution of CTRs generated by the ML model.

To compute $\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right]$ (10), we first perform a change of variables to the logit space:

$$\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] = \frac{\int \sigma(\xi_t) p\left(\hat{\xi}_t \middle| \xi_t\right) p(\xi_t) \mathrm{d}\xi_t}{\int p\left(\hat{\xi}_t \middle| \xi_t\right) p(\xi_t) \mathrm{d}\xi_t}. \tag{15}$$

Next, we substitute the likelihood (13) and the prior (14) into conditional expectation (15):

$$\begin{aligned}&\mathbb{E}\big[CTR_t \big| \widehat{CTR}_t\big] = \\ &= \frac{\sum_{k=1}^{K} \pi_k \int \sigma(\xi_t) \mathcal{N}\big(\hat{\xi}_t \big| \xi_t, \sigma_t^2\big) \mathcal{N}\big(\xi_t \big| \mu_k, \theta_k^2\big) \mathrm{d}\xi_t}{\sum_{k=1}^{K} \pi_k \int \mathcal{N}\big(\hat{\xi}_t \big| \xi_t, \sigma_t^2\big) \mathcal{N}\big(\xi_t \big| \mu_k, \theta_k^2\big) \mathrm{d}\xi_t},\end{aligned} \tag{16}$$

where $\mathcal{N}\left(\cdot \,\middle|\, \mu, \sigma^2\right)$ denotes the Normal Probability Density Function with mean μ and variance σ^2.

By leveraging the property that the product of two Gaussian kernels is a scaled Gaussian and applying the probit approximation to the resulting sigmoid-Gaussian integrals (see Appendix B.1 for the detailed derivation), we obtain a computationally efficient closed-form expression:

$$\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] \approx \sum_{k=1}^{K} \pi'_{t,k} \sigma\left(\frac{\mu'_{t,k}}{\sqrt{1 + \frac{\pi}{8} {\sigma'_{t,k}}^2}}\right), \tag{17}$$

where $\mu'_{t,k}$ and ${\sigma'_{t,k}}^2$ are the posterior mean and variance of the k-th component and $\pi'_{t,k} \propto \pi_k \cdot \mathcal{N}(\hat{\xi}_t \mid \mu_k, \sigma_t^2 + \theta_k^2)$ are the posterior weights.

Empirically reconstructed prior. Implementing the `DenoiseBid` requires the parameters π_k, μ_k, θ_k of the prior distribution $p(\xi)$ introduced in (14). In the real world, we do not have access to ground-truth logits ξ_t, so we can only use noisy estimates $\hat{\xi}_t$ to approximate them.

To recover the prior distribution from noisy samples, we formulate the task as a density estimation problem with measurement errors. Since we assume the prior $p(\xi_t)$ is a Gaussian Mixture, the noise model $p\left(\hat{\xi}_t \mid \xi_t\right)$ is Gaussian, and noise is uncorrelated with CTR, the marginal distribution of the observed logits is also a Gaussian Mixture:

$$\begin{aligned}p(\hat{\xi}_t) &= \int p\big(\hat{\xi}_t \big| \xi_t\big) \cdot p(\xi_t) \mathrm{d}\xi_t = \\ &= \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\hat{\xi}_t \middle| \mu_k, \theta_k^2 + \sigma_t^2\right),\end{aligned} \tag{18}$$

where $\left\{\hat{\xi}_t, \sigma_t^2\right\}$ is CTR model output at auction t.

To estimate the parameters $\boldsymbol{\varphi} = \{\pi_k, \mu_k, \theta_k^2\}_{k=1}^K$, we use the Extreme Deconvolution technique (XDGMM) Bovy et al. [2011] that maximizes the likelihood:

$$\begin{aligned}\mathcal{L}\Big(\boldsymbol{\varphi} \Big| \{\hat{\xi}_t\}_{t=1}^T, \{\sigma_t^2\}_{t=1}^T\Big) &= \sum_{t=1}^{T} \ln p\big(\hat{\xi}_t \big| \boldsymbol{\varphi}, \sigma_t^2\big) = \\ &= \sum_{t=1}^{T} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(\hat{\xi}_t \middle| \mu_k, \theta_k^2 + \sigma_t^2\right)\right).\end{aligned} \tag{19}$$

Unlike the standard EM algorithm, which treats observations as exact, XDGMM considers individual noise variances σ_t^2 for each sample. Therefore, XDGMM is well-suited for reconstructing the CTR distribution for autobidding.

4.2 BIDDING UNDER JOINT CTR-CVR UNCERTAINTY

In the previous section, we assumed independence between CTR and CVR to simplify the derivation. In the case of joint uncertainty, which is closer to real-world scenarios, we consider correlations between CTR and CVR. To capture this relationship, we generalize the proposed framework to the *joint* logit-space $\boldsymbol{\eta}_t = \begin{pmatrix}\xi_t & \zeta_t\end{pmatrix}^\top$, where ξ_t and ζ_t denotes logits of CTR and CVR, respectively:

$$\xi_t = \ln \frac{CTR_t}{1 - CTR_t}, \quad \zeta_t = \ln \frac{CVR_t}{1 - CVR_t}. \tag{20}$$

In this setting, while we still assume independence between CTR and CVR across distinct auctions, we treat them as potentially correlated within each auction. Consequently, the expectation of their product in (6) can no longer be factorized and requires the computing of:

$$\begin{aligned}\mathbb{E}\big[CTR_t \big| \hat{\boldsymbol{\eta}}_t\big] &= \textstyle\int \sigma(\xi_t) p(\boldsymbol{\eta}_t | \hat{\boldsymbol{\eta}}_t) \mathrm{d}^2\boldsymbol{\eta}_t, \\ \mathbb{E}\big[CTR_t CVR_t \big| \hat{\boldsymbol{\eta}}_t\big] &= \textstyle\int \sigma(\xi_t) \sigma(\zeta_t) p(\boldsymbol{\eta}_t | \hat{\boldsymbol{\eta}}_t) \mathrm{d}^2\boldsymbol{\eta}_t,\end{aligned} \tag{21}$$

where $\hat{\boldsymbol{\eta}}_t = \begin{pmatrix}\hat{\xi}_t & \hat{\zeta}_t\end{pmatrix}^\top$ stores logits of noisy CTR and CVR.

Similarly to the one-dimensional case, we express expectations in terms of the likelihood and the prior:

$$\mathbb{E}\big[X\big|\hat{\boldsymbol{\eta}}_t\big] = \frac{\int X \cdot p\big(\hat{\boldsymbol{\eta}}_t\big|\boldsymbol{\eta}_t\big)p(\boldsymbol{\eta}_t)\,\mathrm{d}^2\boldsymbol{\eta}_t}{\int p\big(\hat{\boldsymbol{\eta}}_t\big|\boldsymbol{\eta}_t\big)p(\boldsymbol{\eta}_t)\mathrm{d}^2\boldsymbol{\eta}_t}, \tag{22}$$

where latent variable X denotes CTR_t or $CTR_t \cdot CVR_t$.

Noise model. Following the one-dimensional case, we assume that the observation errors for both CTR and CVR are additive and Gaussian in the logit space:

$$\hat{\boldsymbol{\eta}}_t = \boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t), \tag{23}$$

where $\{\boldsymbol{\varepsilon}_t\}_{t=1}^T$ are independent two-dimensional random vectors and they are independent from $\{\boldsymbol{\eta}_t\}_{t=1}^T$. Then likelihood can be written as:

$$p\big(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\eta}_t\big) = \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_t|}} e^{-\frac{1}{2}(\hat{\boldsymbol{\eta}}_t-\boldsymbol{\eta}_t)^\top\boldsymbol{\Sigma}_t^{-1}(\hat{\boldsymbol{\eta}}_t-\boldsymbol{\eta}_t)}. \tag{24}$$

Gaussian mixture prior. To capture the correlations between CTR and CVR, we model the joint prior distribution of the logit-vector as a bivariate Gaussian Mixture:

$$\begin{gathered}\boldsymbol{\eta}_t \sim \sum_{k=1}^K \pi_k \mathcal{N}\left(\boldsymbol{\mu}_k, \boldsymbol{\Theta}_k\right), \\ p(\boldsymbol{\eta}_t) = \sum_{k=1}^K \frac{\pi_k}{2\pi\sqrt{|\boldsymbol{\Theta}_k|}} e^{-\frac{1}{2}(\boldsymbol{\eta}_t-\boldsymbol{\mu}_k)^\top\boldsymbol{\Theta}_k^{-1}(\boldsymbol{\eta}_t-\boldsymbol{\mu}_k)},\end{gathered} \tag{25}$$

where K is the number of mixture components, π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Theta}_k$ are the weight, mean vector, and covariance matrix of the k-th two-dimensional Gaussian.

Substituting the likelihood (24) and prior (25) into (22) yields the following expression for expectation:

$$\begin{gathered}\mathbb{E}\big[X\big|\hat{\boldsymbol{\eta}}_t\big] = \\ = \frac{\sum_{k=1}^K \pi_k \int X \mathcal{N}(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\eta}_t, \boldsymbol{\Sigma}_t)\mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k)\mathrm{d}^2\boldsymbol{\eta}_t}{\sum_{k=1}^K \pi_k \int \mathcal{N}(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\eta}_t, \boldsymbol{\Sigma}_t)\mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k)\mathrm{d}^2\boldsymbol{\eta}_t}\end{gathered} \tag{26}$$

Similar to the previous case, we can apply the Gaussian distribution property and arrive at the following formula:

$$\mathbb{E}\big[X \mid \hat{\boldsymbol{\eta}}_t\big] = \sum_{k=1}^K \pi'_{t,k} \int X \cdot \mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k})\,\mathrm{d}^2\boldsymbol{\eta}_t, \tag{27}$$

where $\pi'_{t,k}$ are the posterior weights and $\mu'_{t,k}$, $\Sigma'_{t,k}$ are the updated posterior mean vectors and covariance matrices for each component (see detailed derivation in Appendix B.2).

To compute the CTR expectation, we exploit the property that the marginal posterior of ξ_t remains normal and use the probit approximation for each component:

$$\mathbb{E}\big[CTR_t\big|\hat{\boldsymbol{\eta}}_t\big] \approx \sum_{k=1}^K \pi'_{t,k}\sigma\left(\frac{(\boldsymbol{\mu}'_{t,k})_0}{\sqrt{1+\frac{\pi}{8}(\boldsymbol{\Sigma}'_{t,k})_{0,0}}}\right), \tag{28}$$

where $(\cdot)_0$ and $(\cdot)_{0,0}$ denote the first element of the mean vector and the top-left element of the covariance matrix.

However, the value ($X = \sigma(\xi_t)\sigma(\zeta_t)$) involves a product of two sigmoids over a correlated distribution. We employ Gauss-Hermite quadrature Abramowitz and Stegun [1948], Jäckel [2005] to compute the expectation. To consider the correlations, we use the Cholesky decomposition $\boldsymbol{\Sigma}'_{t,k} = \mathbf{L}\mathbf{L}^\top$ to transform the standard nodes to match the mean and covariance of each component:

$$\begin{gathered}\mathbb{E}\big[CTR_tCVR_t\big|\hat{\boldsymbol{\eta}}_t\big] \approx \\ \approx \sum_{k=1}^K \pi'_{t,k} \sum_{i,j=1}^M \lambda_i\lambda_j\sigma\left(\xi_{t,k}^{(i,j)}\right)\sigma\left(\zeta_{t,k}^{(i,j)}\right),\end{gathered} \tag{29}$$

where nodes $\boldsymbol{\eta}_{t,k}^{(i,j)} = \begin{pmatrix}\xi_{t,k}^{(i,j)} & \zeta_{t,k}^{(i,j)}\end{pmatrix}^\top$ are obtained via the Cholesky decomposition of the posterior covariance matrix $\boldsymbol{\Sigma}'_{t,k} = \mathbf{L}_{t,k}\mathbf{L}_{t,k}^\top$ (see details in Appendix B.3).

Since the sigmoid function is smooth, a small number of nodes (e.g., $M = 5$) provides sufficient precision for bidding purposes. Empirically, further increasing M yields no significant accuracy gains. The posterior covariance matrix is only 2×2, so this Cholesky decomposition can be computed analytically. Thus, this approximation scheme is fast and satisfies the latency limits of a real-time bidding system.

Empirically reconstructed prior. As before, the prior parameters $\boldsymbol{\psi} = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$ introduced in (25) must be estimated from the noisy observations $\hat{\boldsymbol{\eta}}_t$. To perform this estimation, we use the property that the marginal distribution of the observed logit vectors is a bivariate Gaussian Mixture:

$$\hat{\boldsymbol{\eta}}_t \sim \sum_{k=1}^K \pi_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Theta}_k + \boldsymbol{\Sigma}_t). \tag{30}$$

where $\boldsymbol{\Sigma}_t$ is the noise covariance provided by the CTR and CVR models (typically diagonal, though `DenoiseBid` can also incorporate error correlations). We infer $\boldsymbol{\psi}$ by maximizing the log-likelihood via the XDGMM Bovy et al. [2011]:

$$\begin{gathered}\mathcal{L}(\boldsymbol{\psi} \mid \{\hat{\boldsymbol{\eta}}_t\}_{t=1}^T, \{\boldsymbol{\Sigma}_t\}_{t=1}^T) = \sum_{t=1}^T \ln p\left(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\psi}, \boldsymbol{\Sigma}_t\right) = \\ = \sum_{t=1}^T \ln\left(\sum_{k=1}^K \pi_k \mathcal{N}\left(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k + \boldsymbol{\Sigma}_t\right)\right).\end{gathered} \tag{31}$$

Note that XDGMM can work in parallel or in the background regime to avoid redundant overhead. Therefore, it is tractable for deployment in the real-time bidding system.

5 COMPUTATIONAL EXPERIMENT

In this section, we evaluate the `DenoiseBid` framework against two baselines: the non-robust strategy Yang et al.

[2019] and the Robust Optimization approach Pudovikov et al. [2026]. We use `DenoiseBid` under two prior distributions in the logit space: a normal distribution and a two-component Gaussian Mixture. For noise modeling, we use two settings: synthetic noise and empirical noise extracted from the pre-trained CTR model. To reproduce the presented results, the source code is available at `https://github.com/hxrxmx/denoise-bid-uai26/`.

Setup. We use an offline setting to compare the considered algorithms. To ensure stable behavior patterns, we use relative budget and CPC constraints defined by fixed scaling factors k_B and k_C across campaigns:

$$B = k_B \sum_{t=1}^{T} wp_t, \quad C = k_C \frac{\sum_{t=1}^{T} wp_t}{\sum_{t=1}^{T} CTR_t}. \tag{32}$$

Further, we use $k_B = 0.2$ and $k_C = 0.2$ to define budget and CPC upper bounds.

Datasets. First, we use fully-synthetic data to verify the consistency of our CTR-only and joint methods. Then, we move to real-world cases using iPinYou and BAT datasets with empirical winning prices and model predictions. Finally, we test `DenoiseBid` on the Criteo Attribution dataset. We train a CatBoost CTR model on this dataset and use CatBoost Virtual Ensembles Malinin et al. [2021] to estimate knowledge uncertainty in CTR values. This setup does not include any sampling of synthetic noise or data.

Table 2: Summary of experimental settings across datasets. The red cross indicates that the dataset admits such a type of noise; however, we skip this setting due to limited resources. A dash indicates that the noise type is infeasible.

Dataset	Synthetic noise	Empirical noise
Synthetic	✓	—
iPinYou	✓	✗
BAT	✓	—
Criteo Attribution	✗	✓

5.1 EXPERIMENTS WITH SYNTHETIC NOISE

These experiments verify the mathematical consistency of our methods using known noise structures.

Datasets and setup. We evaluate our approach on three datasets. In the **Synthetic** dataset, winning prices are generated from a log-normal distribution, and CTR/CVR logits are generated from a Gaussian mixture. For **iPinYou**, we use empirical winning prices, and CTR/CVR logits are generated from a Gaussian mixture. Finally, the **BAT** dataset leverages both real-world winning prices and model predictions of CTR/CVR directly from historical auction logs. Details of the experiment are provided in the Appendix C.1.

Evaluation metrics. To evaluate performance and constraint adherence, we define two key indicators:

- The ratio of the optimal value for objective functions in problem (3) to the theoretical optimum from problem (1) denoted as R/R^*.
- $\overline{CPC}/\overline{CPC}_{\text{camp}}$: the actual average cost-per-click normalized by the campaign's mean CPC. Since the target limit is set as $C = k_C \cdot \overline{CPC}_{\text{camp}}$, a strategy is considered cost-compliant only if this ratio is below the threshold k_C.

Accuracy of the first-order approximation. To study the accuracy of the first-order approximation used in (4), we empirically compare the surrogate CPC against the value estimated by Monte-Carlo sampling on the BAT dataset. Figure 1 shows the relative error across a wide range of noise in CTR logits. The relative error remains below 7% even with extreme noise levels. Moreover, the surrogate consistently exceeds the MC sampling value, implying that DenoiseBid yields cheaper clicks in the high-noise regime.

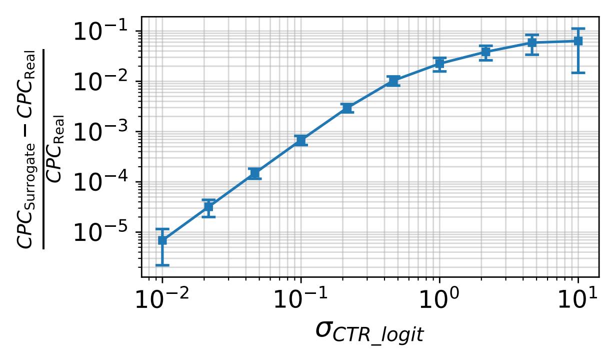

Figure 1: The relative error between the surrogate value and Monte Carlo estimates remains negligible for typical noise scales, which confirms the correctness of estimation in (4).

Main results for CTR-only uncertainty. Figure 2 illustrates the behavior of the strategies under increasing noise σ^{CTR}. The non-robust baseline fails as noise increases, violating CPC limits and reducing efficiency. While `RobustBid` remains CPC-compliant, it drops R/R^*. An additional baseline `RiskBid` regularizes the non-robust objective by using a standard deviation of CTR weighted by a hyperparameter α. We consider α equal to 0.1 and 0.5. `RiskBid` violates the CPC constraint even in the low-noise regime. In contrast, `DenoiseBid` demonstrates the highest stability as noise increases, providing the best trade-off between conversion volume and constraint adherence.

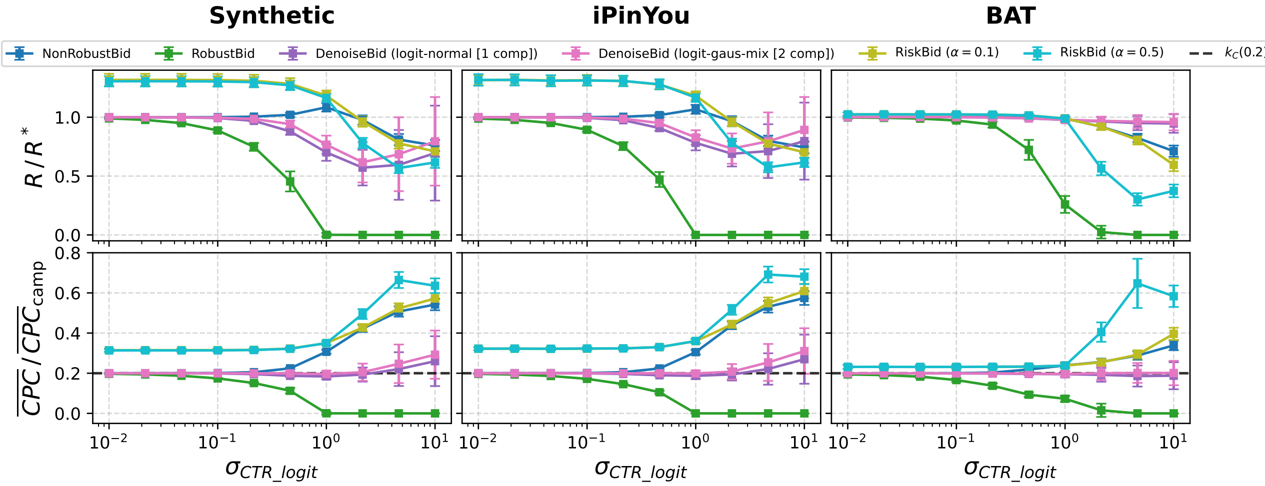

Figure 2: Performance comparison on the synthetic, iPinYou and BAT datasets under CTR uncertainty and synthetic noise.

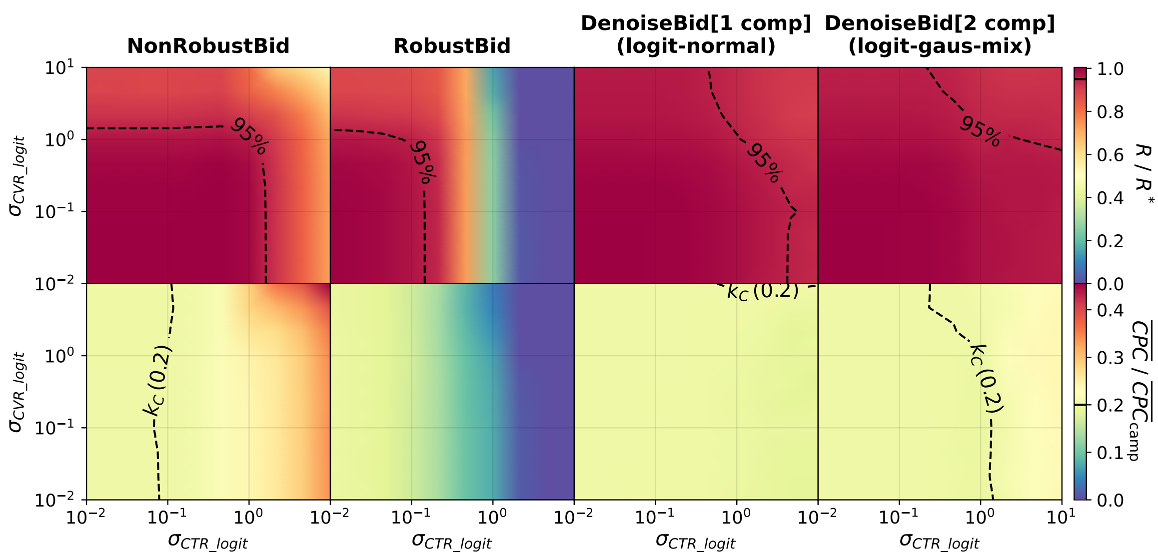

Figure 3: Performance comparison on the BAT dataset under joint CTR-CVR uncertainty and synthetic noise.

Main results for joint CTR-CVR uncertainty. Figure 3 illustrates R/R^* and $\overline{CPC}/\overline{CPC}_{\text{camp}}$ heatmaps under bivariate noise for the BAT dataset. The non-robust strategy fails to maintain cost compliance over the noise range, whereas `RobustBid` achieves compliance at the cost of a significantly lower R/R^*. In contrast, `DenoiseBid` maintains a stable, near-optimal conversion volume while almost strictly adhering to constraints. We have tested the single-component logit-normal prior and two-component logit-gaus-mix prior. These priors yield nearly the same results, whereas the single-component logit-normal prior leads to cheaper clicks. At the same time, the two-component logit-gaus-mix prior gives a slightly higher conversion ratio. Similar heatmaps for the Synthetic and iPinYou datasets are presented in Appendix D and show similar behavior.

What if CTR and CVR are independent? This paragraph examines the simplified case in which CTR and CVR are independent. Then, the expectation of their product factorizes into a product of individual expectations:

$$\begin{aligned} \mathbb{E}\left[CTR_t\, CVR_t | \widehat{CTR}_t, \widehat{CVR}_t\right] &\approx \\ \approx \mathbb{E}\left[CTR_t | \widehat{CTR}_t\right] &\cdot \mathbb{E}\left[CVR_t | \widehat{CVR}_t\right]. \end{aligned} \tag{33}$$

This approximation is computationally efficient since it yields a closed-form bidding rule:

$$\begin{aligned} bid_t =& \frac{1}{p^* + q^*}\mathbb{E}[CTR_t|\widehat{CTR}_t] \cdot \mathbb{E}[CVR_t|\widehat{CVR}_t] + \\ &+ \frac{q^*}{p^* + q^*} C \cdot \mathbb{E}[CTR_t \mid \widehat{CTR}_t], \end{aligned} \tag{34}$$

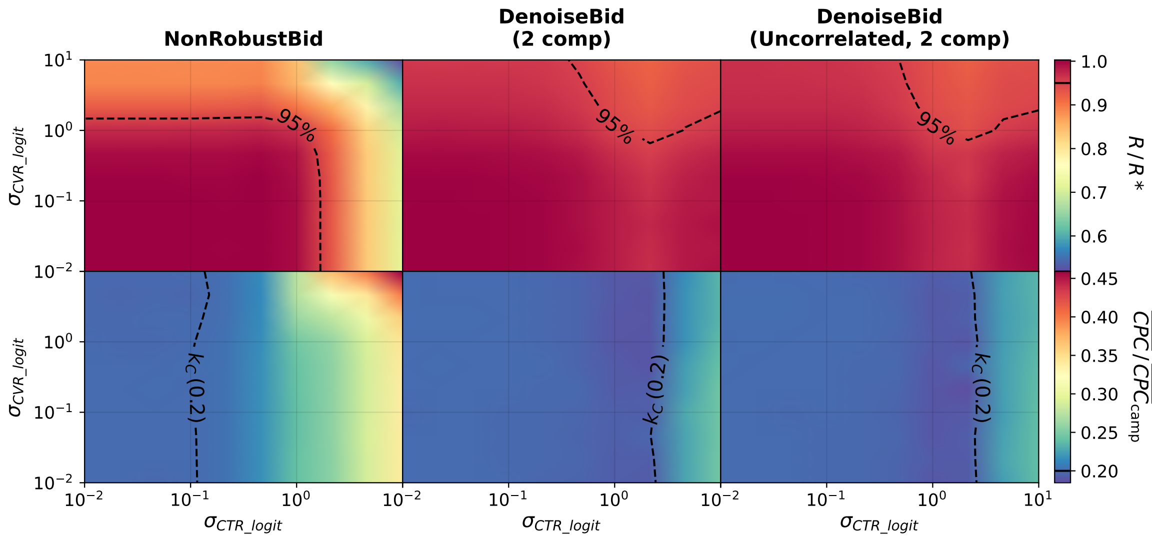

Figure 4: Performance comparison on the BAT dataset under joint uncertainty. The top row shows expected conversion ratio (R/R^{*}), and the bottom row shows normalized CPC deviation. Both `DenoiseBid` variants maintain high conversion volume and cost compliance, whereas the uncorrelated variant offers a faster alternative with negligible loss in accuracy.

where each expectation can be calculated using the 1D probit approximation (28), with complete analogy to the CTR-only uncertainty case. Thus, one does not need to use the multivariate numerical integration techniques.

Figure 4 demonstrates that the uncorrelated variant of `DenoiseBid` is robust to noise nearly identical to the full joint model. It maintains a stable, near-optimal conversion volume and adheres almost strictly to the CPC constraint, even under extreme noise. These observations suggest that the primary driver of `DenoiseBid` performance is the Bayesian treatment of individual uncertainties, while the impact of CTR-CVR correlations is secondary. Thus, an uncorrelated variant of `DenoiseBid` serves as a high-performance alternative for latency-critical systems.

Latency analysis. To evaluate the feasibility of `DenoiseBid` in real-time bidding (RTB) environments, we measured the average computation time per bid (see Table 3). While our Bayesian approach introduces additional overhead compared to the baseline, the latency remains under $200\mu s$ even for the most complex configuration (joint algorithm with 5×5 nodes and a 3-component GMM prior). Therefore, `DenoiseBid` remains feasible for deployment.

Table 3: Average bidding latency per auction indicates that `DenoiseBid` still remains feasible for deployment although it leads to an increase in runtime. Number of components refers to the number of Gaussians in the prior; $M \times M$ nodes denotes the number of quadrature points used for Gauss-Hermite integration in the joint uncertainty case.

Method	**Runtime**, μs	**Std**, μs
`NonRobustBid`	3.2	1.4
`RobustBid`	3.1	0.36
`DenoiseBid` CTR-only		
1 component	13.8	1.3
2 component	13.8	0.1
3 component	13.8	0.2
`DenoiseBid` Joint, 3×3 nodes		
1 component	156.6	10.3
2 component	175.7	3.1
3 component	198.2	7.2
`DenoiseBid` Joint, 5×5 nodes		
1 component	158.6	14.5
2 component	175.8	4.3
3 component	197.5	5.7

5.2 EXPERIMENTS WITH EMPIRICAL NOISE

This section uses the Criteo Attribution logs to evaluate `DenoiseBid` using uncertainties inferred from the trained CTR model. We estimate predictive noise via Virtual Ensembles Malinin et al. [2021], which, unlike standard ensembles, provide per-sample logit variances without training multiple independent models. These variances are plugged directly into `DenoiseBid` as diagonal entries of covariance matrices $\mathbf{\Sigma}_t$. In contrast to the synthetic noise setup, we cannot directly affect these variances. Therefore, we have considered two proxy approaches to vary noise levels:

1. **Feature removal**, where features are excluded based on their importance, which is a proxy for uncertainty

2. **Data scaling**, where training set size reduces (from 100% to 0.1%) to increase knowledge uncertainty.

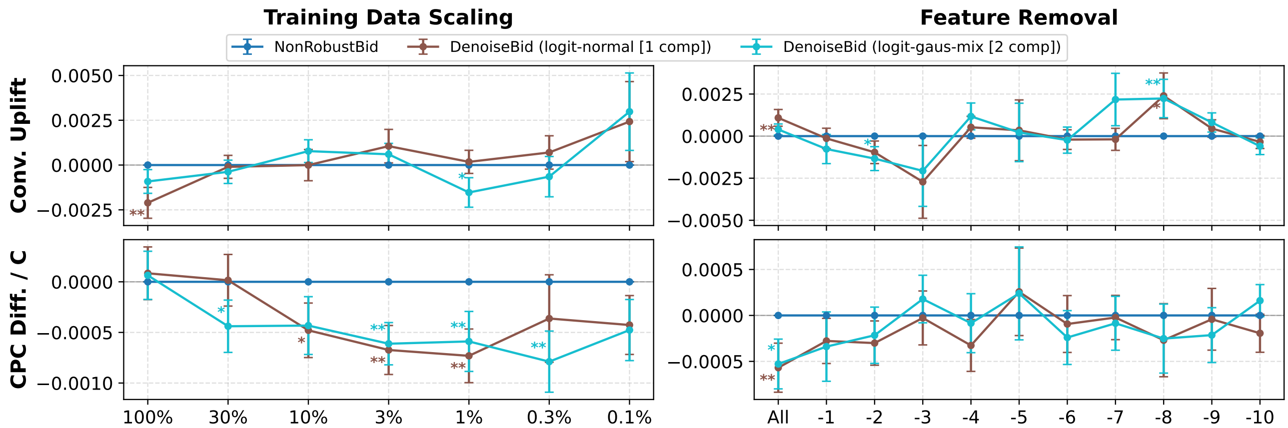

Figure 5: Relative uplifts in conversions and CPC deviation on the Criteo Attribution dataset. Markers * and ** denote statistical significance at the 10% and 5% levels, respectively.

Details of these procedures are presented in Appendix C.2.

In contrast to the synthetic case, we evaluate uplifts using logged conversions and clicks from the dataset. Although we cannot model market feedback or competitor reactions to our bidding strategy, this approach avoids model-induced prediction errors in target metrics. Therefore, we formalize the conversion- and CPC-related analogs of the target metrics in the following way:

- relative error $(conv - conv_{nr})/conv_{nr}$, where $conv$ denotes the number of conversions for `DenoiseBid` and $conv_{nr}$ corresponds to the non-robust baseline.
- $(\overline{CPC} - \overline{CPC}_{nr})/C$, representing the cost change normalized by the target C.

Thus, the evaluation based on the observed conversions rather than predicted CVRs effectively decouples the autobidding performance from imperfect model's errors.

Main results. Figure 5 summarizes the relative uplifts. The logit-normal configuration yields a statistically significant ($p < 0.05$) reduction in CPC deviation across the 1% and 3% data-scaling regimes. The logit-gaus-mix variant with two components in the GMM prior demonstrates broader cost stability, yielding significant CPC improvements across the 0.3%, 1%, and 3% data regimes. Furthermore, it achieves significant conversion gains in the feature removal scenario when the 8 least-informative features are excluded. The plots include error bars representing the standard error of the mean to demonstrate the statistical significance of the observed performance gains. Note that the presented plots do not show the clear trend in dependence of conversion and CPC on the considered noise-inducing factors. A potential reason for this effect is the limitations of virtual ensembles in predicting noise in our application. The alternative approaches require further investigation and can be easily incorporated into the `DenoiseBid` framework.

6 LIMITATIONS AND FUTURE WORK

We provide experiments only with CatBoost models for conversion prediction; however, they are likely to become less popular as deep learning approaches become increasingly widespread. Therefore, we plan to adapt state-of-the-art models from Zhu et al. [2022] for our `DenoiseBid` framework. The framework will remain the same; the main difference is in the specific steps for handling uncertainty.

7 CONCLUSION

We have proposed `DenoiseBid`, a Bayesian autobidding method that explicitly accounts for uncertainty in CTR and CVR estimates. The core idea is to replace the deterministic LP-based bidding formulation with a stochastic optimization problem in which bids are computed from posterior expectations of click and conversion probabilities. Since the Bayesian treatment of noisy CTR and CVR values requires prior distributions over the ground-truth CTR and CVR, we consider a Gaussian Mixture Model as the prior. Then, the Extreme Deconvolution algorithm yields the appropriate parameters for the GMM. To evaluate the performance of our approach, we conducted extensive experiments on four datasets: Synthetic, iPinYou, BAT, and Criteo Attribution. The experiments with the synthetic noise cover the Synthetic, iPinYou, and BAT datasets and show that `DenoiseBid` maintains near-optimal conversion volume while satisfying CPC constraints, outperforming both the non-robust baseline and the `RobustBid` strategy across a wide range of noise levels. In addition, we use the Criteo Attribution dataset to test `DenoiseBid` under empirical noise from the trained CTR and CVR models. In this setting, `DenoiseBid` achieves a higher conversion rate and a lower CPC compared to a non-robust baseline.

References

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information fusion*, 76:243–297, 2021.

Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office, 1948.

Gagan Aggarwal, Ashwinkumar Badanidiyuru, Santiago R Balseiro, Kshipra Bhawalkar, Yuan Deng, Zhe Feng, Gagan Goel, Christopher Liaw, Haihao Lu, Mohammad Mahdian, et al. Auto-bidding and auctions in online advertising: A survey. *ACM SIGecom Exchanges*, 22(1): 159–183, 2024.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Jo Bovy, David W Hogg, and Sam T Roweis. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *The Annals of Applied Statistics*, 5(2B):1657, 2011.

Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. Real-Time Bidding by Reinforcement Learning in Display Advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 661–670, 2017.

Yue He, Xiujun Chen, Di Wu, Junwei Pan, Qing Tan, Chuan Yu, Jian Xu, and Xiaoqiang Zhu. A Unified Solution to Constrained Bidding in Online Display Advertising. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 2993–3001, 2021.

Peter Jäckel. A note on multivariate Gauss-Hermite quadrature. *London: ABN-Amro. Re*, 2005.

Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5580–5590, 2017.

Alexandra Khirianova, Ekaterina Solodneva, Andrey Pudovikov, Sergey Osokin, Egor Samosvat, Yuriy Dorn, Alexander Ledovsky, and Yana Zenkova. BAT: Benchmark for Auto-bidding Task. In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 2657–2667, 2025.

Andrey Malinin, Liudmila Prokhorenkova, and Aleksei Ustimenko. Uncertainty in Gradient Boosting via Ensembles. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=1Jv6b0Zq3qi`.

Weitong Ou, Bo Chen, Xinyi Dai, Weinan Zhang, Weiwen Liu, Ruiming Tang, and Yong Yu. A survey on bid optimization in real-time bidding display advertising. *ACM Transactions on Knowledge Discovery from Data*, 18(3): 1–31, 2023.

Andrey Pudovikov, Alexandra Khirianova, Ekaterina Solodneva, Aleksandr Katrutsa, Egor Samosvat, and Yuriy Dorn. Autobidding Arena: unified evaluation of the classical and RL-based autobidding algorithms. *arXiv preprint arXiv:2510.19357*, 2025.

Andrey Pudovikov, Khirianova Alexandra, Ekaterina Solodneva, Gleb Molodtsov, Aleksandr Katrutsa, Yuriy Dorn, and Egor Samosvat. Robust Autobidding for Noisy Conversion Prediction Models. In *Proceedings of the 25th International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '26, page 1379–1387, 2026.

Wen-Yueh Shih, Hsu-Chao Lai, and Jiun-Long Huang. A Robust Real Time Bidding Strategy Against Inaccurate CTR Predictions by Using Cluster Expected Win Rate. *IEEE Access*, 11:126917–126926, 2023.

Rotem Stram, Rani Abboud, Alex Shtoff, Oren Somekh, Ariel Raviv, and Yair Koren. Mystique: A Budget Pacing System for Performance Optimization in Online Advertising. In *Companion Proceedings of the ACM Web Conference 2024*, WWW '24, page 433–442, 2024.

Xun Yang, Yasong Li, Hao Wang, Di Wu, Qing Tan, Jian Xu, and Kun Gai. Bid Optimization by Multivariable Control in Display Advertising. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, page 1966–1974, 2019.

Haifeng Zhang, Weinan Zhang, Yifei Rong, Kan Ren, Wenxin Li, and Jun Wang. Managing Risk of Bidding in Display Advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 581–590, 2017.

Jieming Zhu, Quanyu Dai, Liangcai Su, Rong Ma, Jinyang Liu, Guohao Cai, Xi Xiao, and Rui Zhang. BARS: Towards Open Benchmarking for Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 2912–2923, 2022.

Uncertainty Quantification of Click and Conversion Estimates for the Autobidding (Supplementary Material)

Ivan Zhigalskii[3] **Andrey Pudovikov**[1,2] **Aleksandr Katrutsa**[1] **Egor Samosvat**[4]

[1]AI Center, Lomonosov Moscow State University, Moscow, Russia,
[2]IAI, Lomonosov Moscow State University, Moscow, Russia,
[3]HSE, Moscow, Russia,
[4]Independent researcher, Moscow, Russia,

This Supplementary Material provides derivation of the main formulas and auxiliary details and results of experiments.

A SOLUTION OF A STOCHASTIC OPTIMIZATION PROBLEM

In this section, we provide the detailed derivation of the optimal bidding policy for the stochastic optimization problem. We start with the linearized primal formulation of the `DenoiseBid` problem:

$$\begin{aligned} \max_{x_t} \quad & \sum_{t=1}^{T} x_t v_t, \\ \text{s.t.} \quad & \sum_{t=1}^{T} x_t \cdot wp_t \leq B, \\ & \sum_{t=1}^{T} x_t \cdot (wp_t - C \cdot c_t) \leq 0, \\ & 0 \leq x_t \leq 1, \quad t = 1, \dots, T, \end{aligned} \tag{35}$$

where $v_t = \mathbb{E}[CTR_t \cdot CVR_t \mid \mathcal{O}]$ and $c_t = \mathbb{E}[CTR_t \mid \mathcal{O}]$ are the posterior means of the conversion value and click probability, respectively.

Lagrangian formulation. To solve problem (35), we derive its Lagrangian dual. We introduce non-negative dual variables $p \geq 0$ and $q \geq 0$ associated with the budget and CPC constraints, respectively. To handle the constraints $x_t \leq 1$, we also introduce auxiliary dual variables $r_t \geq 0$ for each auction. The Lagrangian $\mathcal{L}$ is formulated as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, p, q, \mathbf{r}) = \sum_{t=1}^{T} x_t v_t - p \left(\sum_{t=1}^{T} x_t wp_t - B \right) - \\ - q \left(\sum_{t=1}^{T} x_t (wp_t - C \cdot c_t) \right) - \sum_{t=1}^{T} r_t (x_t - 1), \end{aligned} \tag{36}$$

where $c_t = \mathbb{E}\left[CTR_t | \mathcal{O}\right]$, $v_t = \mathbb{E}\left[CTR_t \cdot CVR_t | \mathcal{O}\right]$ denote the posterior means of click probability and value at auction t, respectively. By rearranging terms to group the variables x_t, the dual objective is obtained by minimizing the Lagrangian with respect to $\mathbf{x}$.

$$\begin{aligned} \min_{p,q,r_t} \quad & B \cdot p + \sum_{t=1}^{T} r_t, \\ \text{s.t.} \quad & wp_t \cdot p + (wp_t - C \cdot c_t) \cdot q + r_t \geq v_t, \quad \forall t, \\ & p \geq 0, q \geq 0, r_t \geq 0. \end{aligned} \tag{37}$$

Optimality conditions. According to the Theorem of Complementary Slackness, the optimal primal solutions $\mathbf{x}^*$ and dual solutions p^*, q^* and $\mathbf{r}^*$ must satisfy:

$$x_t^* \cdot (v_t - wp_t \cdot p^* - (wp_t - C \cdot c_t) \cdot q^* - r_t^*) = 0, \tag{38}$$

$$(x_t^* - 1) \cdot r_t^* = 0. \tag{39}$$

We define the optimal bid price bid_t^* as:

$$bid_t^* = \frac{1}{p^* + q^*} v_t + \frac{q^*}{p^* + q^*} C \cdot c_t, \tag{40}$$

By substituting (40) into the condition (38), we obtain:

$$x_t^* \cdot ((bid_t^* - wp_t)(p^* + q^*) - r_t^*) = 0. \tag{41}$$

Then, as it follows from (39) and (41),

- if the bidding agent wins the auction ($x_t^* > 0$), then $(bid_t^* - wp_t)(p^* + q^*) - r_t^* = 0$. Thus, since $p^*, q^*, r_t^* \geq 0$, it must be $bid_t^* \geq wp_t$.
- if bidding agent loses the auction ($x_t^* = 0$), then $r_t^* = 0$. Therefore, from dual condition $wp_t \cdot p + (wp_t - C \cdot c_t) \cdot q \geq v_t$, and it can be rewritten as $bid_t^* \leq wp_t$.

As a result, if p and q are the optimal solution of dual problem (37), optimal bidding formula is:

$$bid_t = \frac{1}{p+q} \mathbb{E}\left[CTR_t \cdot CVR_t | \mathcal{O}\right] + \frac{q}{p+q} C \cdot \mathbb{E}\left[CTR_t | \mathcal{O}\right]. \tag{42}$$

B DERIVATION OF THE CALCULATION FORMULAS FOR EXPECTATIONS

B.1 CTR-ONLY UNCERTAINTY

Reduction of Gaussian products. To simplify the expression:

$$\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] = \frac{\sum_k \pi_k \int \sigma(\xi_t) \cdot \mathcal{N}\left(\hat{\xi}_t \middle| \xi_t, \sigma_t^2\right) \mathcal{N}\left(\xi_t \mid \mu_k, \theta_k^2\right) \, \mathrm{d}\xi_t}{\sum_k \pi_k \int \mathcal{N}\left(\hat{\xi}_t \middle| \xi_t, \sigma_t^2\right) \mathcal{N}\left(\xi_t \mid \mu_k, \theta_k^2\right) \, \mathrm{d}\xi_t}, \tag{43}$$

we leverage a fundamental property of Normal densities: the product of two Gaussian kernels is itself a scaled Gaussian kernel:

$$\mathcal{N}\left(\hat{\xi}_t \middle| \xi_t, \sigma_t^2\right) \cdot \mathcal{N}\left(\xi_t \middle| \mu_k, \theta_k^2\right) = \alpha_{t,k} \cdot \mathcal{N}\left(\xi_t \middle| \mu'_{t,k}, {\sigma'_{t,k}}^2\right), \tag{44}$$

where scaling factor $\alpha_{t,k} = \mathcal{N}\left(\hat{\xi}_t \middle| \mu_k, \sigma_t^2 + \theta_k^2\right)$ represents the evidence of the k-th component, and

$$\begin{aligned} \alpha_{t,k} &= \frac{1}{\sqrt{2\pi\left(\sigma_t^2 + \theta_k^2\right)}} e^{-\frac{(\hat{\xi}_t - \mu_k)^2}{2\left(\sigma_t^2 + \theta_k^2\right)}}, \\ {\sigma'_{t,k}}^2 &= \frac{1}{1/\sigma_t^2 + 1/\theta_k^2}, \\ \mu'_{t,k} &= {\sigma'_{t,k}}^2 \left(\frac{\hat{\xi}_t}{\sigma_t^2} + \frac{\mu_k}{\theta_k^2}\right). \end{aligned} \tag{45}$$

Substituting the Gaussian product property (44) into the expectation (43), we observe that the posterior expectation is a weighted average of sigmoid-Gaussian integrals:

$$\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] = \frac{\sum_k \pi_k \alpha_{t,k} \int \sigma(\xi_t) \mathcal{N}\left(\xi_t \middle| \mu'_{t,k}, {\sigma'_{t,k}}^2\right) \, \mathrm{d}\xi_t}{\sum_k \pi_k \alpha_{t,k}}. \tag{46}$$

Probit approximation. The integral of a sigmoid with a Gaussian density does not have an exact closed-form solution in terms of elementary functions. To obtain a computationally efficient result, we utilize the probit approximation:

$$\int \sigma(\xi)\mathcal{N}\left(\xi \,\middle|\, \mu, \sigma^2\right) \,\mathrm{d}\xi \approx \sigma\left(\frac{\mu}{\sqrt{1+\frac{\pi}{8}\sigma^2}}\right). \tag{47}$$

Applying this approximation to each component in (46), we present the final closed-form expression for the denoised CTR:

$$\mathbb{E}\left[CTR_t \middle| \widehat{CTR}_t\right] \approx \sum_k \pi'_{t,k}\sigma\left(\frac{\mu'_{t,k}}{\sqrt{1+\frac{\pi}{8}{\sigma'_{t,k}}^2}}\right), \tag{48}$$

where $\mu'_{t,k}$ and ${\sigma'_{t,k}}^2$ can be computed via (45), and $\pi'_{t,k}$ denotes posterior probability of k-th Gaussian given by t-th observation:

$$\pi'_{t,k} = \frac{\pi_k \alpha_{t,k}}{\sum_k \pi_k \alpha_{t,k}}. \tag{49}$$

B.2 JOINT UNCERTAINTY

Reduction of bivariate Gaussian Products. We have an expression:

$$\mathbb{E}\left[X \,\middle|\, \hat{\boldsymbol{\eta}}_t\right] = \frac{\sum_k \pi_k \int X \cdot \mathcal{N}(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\eta}_t, \boldsymbol{\Sigma}_t)\mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k)\,\mathrm{d}^2\boldsymbol{\eta}_t}{\sum_k \pi_k \int \mathcal{N}(\hat{\boldsymbol{\eta}}_t \mid \boldsymbol{\eta}_t, \boldsymbol{\Sigma}_t)\mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k)\,\mathrm{d}^2\boldsymbol{\eta}_t}. \tag{50}$$

By complete analogy with one-dimensional Gaussians (44), the product of normal densities is equal to normal density with scaling factor:

$$\mathcal{N}\left(\hat{\boldsymbol{\eta}}_t \,\middle|\, \boldsymbol{\eta}_t, \boldsymbol{\Sigma}_t\right) \cdot \mathcal{N}\left(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Theta}_k\right) = \alpha_{t,k} \cdot \mathcal{N}\left(\boldsymbol{\eta}_t \,\middle|\, \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k}\right), \tag{51}$$

where $\alpha_{t,k} = \mathcal{N}\left(\hat{\boldsymbol{\eta}}_t \,\middle|\, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_t + \boldsymbol{\Theta}_k\right)$ is the evidence of the k-th Gaussian, and

$$\begin{aligned}
\alpha_{t,k} &= \frac{e^{-\frac{1}{2}(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\mu}_k)^\top(\boldsymbol{\Sigma}_t + \boldsymbol{\Theta}_k)^{-1}(\hat{\boldsymbol{\eta}}_t - \boldsymbol{\mu}_k)}}{2\pi\sqrt{|\boldsymbol{\Sigma}_t + \boldsymbol{\Theta}_k|}}, \\
\boldsymbol{\Sigma}'_{t,k} &= \left(\boldsymbol{\Sigma}_t^{-1} + \boldsymbol{\Theta}_k^{-1}\right)^{-1}, \\
\boldsymbol{\mu}'_{t,k} &= \boldsymbol{\Sigma}'_{t,k}\left(\boldsymbol{\Sigma}_t^{-1}\hat{\boldsymbol{\eta}}_t + \boldsymbol{\Theta}_k^{-1}\boldsymbol{\mu}_k\right).
\end{aligned} \tag{52}$$

Next, we substitute simplified form of normal densities product (51) in posterior expectation expression (50):

$$\mathbb{E}\left[X \,\middle|\, \hat{\boldsymbol{\eta}}_t\right] = \frac{\sum_k \pi_k \alpha_{t,k} \int X \cdot \mathcal{N}\left(\boldsymbol{\eta}_t \,\middle|\, \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k}\right)\,\mathrm{d}^2\boldsymbol{\eta}_t}{\sum_k \pi_k \alpha_{t,k}}, \tag{53}$$

or, in closer form, with redefining posterior probability $\pi'_{t,k} = \frac{\pi_k \alpha_{t,k}}{\sum_k \pi_k \alpha_{t,k}}$ of k-th Gaussian given by t-th estimates:

$$\mathbb{E}\left[X \,\middle|\, \hat{\boldsymbol{\eta}}_t\right] = \sum_k \pi'_{t,k} \int X \cdot \mathcal{N}\left(\boldsymbol{\eta}_t \,\middle|\, \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k}\right)\,\mathrm{d}^2\boldsymbol{\eta}_t. \tag{54}$$

B.3 NUMERICAL APPROXIMATION OF INTEGRAL

Probit approximation for CTR expectation. Now, we focus on approximating the integrals in (54). For case with the click-through rate expectation $X = CTR_t = \sigma(\xi_t)$, the integral involves only a single sigmoid function over a bivariate normal distribution. Since the marginal distribution of ξ_t is still Gaussian, we can directly apply the probit approximation as in the one-dimensional case:

$$\mathbb{E}\left[CTR_t \,\middle|\, \hat{\boldsymbol{\eta}}_t\right] \approx \sum_k \pi'_{t,k}\sigma\left(\frac{\left(\boldsymbol{\mu}'_{t,k}\right)_0}{\sqrt{1+\frac{\pi}{8}\left(\boldsymbol{\Sigma}'_{t,k}\right)_{0,0}}}\right), \tag{55}$$

where $(\,\cdot\,)_0$ and $(\,\cdot\,)_{0,0}$ denote the first element of vector and the top left element of matrix, respectively.

Numerical approximation via Gauss-Hermite quadrature for value expectation. The expectation of the conversion value, when $X = CTR_t \cdot CVR_t = \sigma(\xi_t) \cdot \sigma(\zeta_t)$, involves the product of two sigmoid functions. Probit approximation has no multi-dimensional generalization, so, to calculate it with high computational efficiency while preserving accuracy, we employ Gauss-Hermite quadrature.

This method approximates convolution of function $f(\boldsymbol{\eta})$ with normal kernel by evaluating it at $M \times M$ pre-defined nodes:

$$\int f(\boldsymbol{\eta}) \cdot \mathcal{N}(\boldsymbol{\eta} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \mathrm{d}^2\boldsymbol{\eta} \approx \sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j f\left(\boldsymbol{\eta}^{(i,j)}\right), \tag{56}$$

where the nodes $\boldsymbol{\eta}^{(i,j)}$ are obtained via the Cholesky decomposition of the posterior covariance matrix $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^\top$ as:

$$\boldsymbol{\eta}^{(i,j)} = \boldsymbol{\mu} + \sqrt{2}\mathbf{L} \begin{pmatrix} a_i \\ a_j \end{pmatrix}, \tag{57}$$

and $\{a_i\}_i$, $\{\lambda_i\}_i$ are the standard Gauss-Hermite nodes and weights.

Since the sigmoid function is smooth, a small number of nodes (e.g., $M = 5$) provides sufficient precision for bidding purposes. The posterior covariance matrix is only 2×2, so this Cholesky decomposition can be computed analytically. Thus, this approximation scheme is highly efficient, satisfying the latency limits of the Real-Time Bidding task. Therefore, for each auction t separately, we can approximate:

$$\int \sigma(\xi_t)\sigma(\zeta_t) \cdot \mathcal{N}(\boldsymbol{\eta}_t \mid \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k}) \, \mathrm{d}^2\boldsymbol{\eta}_t \approx \sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j \cdot \sigma\left(\xi_{t,k}^{(i,j)}\right) \sigma\left(\zeta_{t,k}^{(i,j)}\right), \tag{58}$$

Calculation formula. As a result, we can approximate every integral in sum (54) and provide the final expression:

$$\mathbb{E}[CTR_t \cdot CVR_t \mid \boldsymbol{\eta}_t] \approx \sum_k \pi'_{t,k} \sum_{i=1}^{M} \sum_{j=1}^{M} \lambda_i \lambda_j \cdot \sigma\left(\xi_{t,k}^{(i,j)}\right) \sigma\left(\zeta_{t,k}^{(i,j)}\right), \tag{59}$$

where $\pi'_{t,k} = \frac{\pi_k \alpha_{t,k}}{\sum_k \pi_k \alpha_{t,k}}$,

$$\begin{aligned} \begin{pmatrix} \xi_{t,k}^{(i,j)} \\ \zeta_{t,k}^{(i,j)} \end{pmatrix} &= \boldsymbol{\mu}'_{t,k} + \sqrt{2}\mathbf{L}_{t,k} \begin{pmatrix} a_i \\ a_j \end{pmatrix}, \\ \mathbf{L}_{t,k} &= \begin{pmatrix} \sqrt{s_1} & 0 \\ s_0/\sqrt{s_1} & \sqrt{s_2 - s_0^2/s_1} \end{pmatrix}, \\ \text{where } \begin{pmatrix} s_1 & s_0 \\ s_0 & s_2 \end{pmatrix} &= \boldsymbol{\Sigma}'_{t,k}. \end{aligned} \tag{60}$$

Parameters $\alpha_{t,k}, \boldsymbol{\mu}'_{t,k}, \boldsymbol{\Sigma}'_{t,k}$ defined in (45), and a_k, λ_k are tabular Gauss-Hermite nodes and weights.

C EXPERIMENT DETAILS

In this section, we provide complete details for the datasets, model architectures, and hyperparameters used in our evaluation.

As was mentioned in Section 5, we utilize the relative constraints:

$$\begin{aligned} B &= k_B \cdot \sum_t wp_t, \\ C &= k_C \cdot \frac{\sum_t wp_t}{\sum_t CTR_t}. \end{aligned} \tag{61}$$

In all experiments, $k_C = 0.2$ and $k_B = 0.2$ were used.

C.1 EXPERIMENTS WITH SYNTHETIC NOISE (SYNTHETIC, IPINYOU, BAT)

Synthetic

- Number of auctions $T = 1000$, number of campaigns $N = 200$.
- winning price wp sampled from a log-normal distribution $\ln wp \sim \mathcal{N}(\mu_{wp}, \sigma_{wp}^2)$ with $\mu_{wp} = \ln 100.0 - \sigma_{wp}$, $\sigma_{wp} = 1.0$.
- CTR sampled from a log-normal distribution $\operatorname{logit} CTR \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \sigma_k^2)$ with $K = 3$, $\{\pi_1 = 0.6, \pi_2 = 0.2, \pi_3 = 0.2\}$, $\{\mu_1 = -3.0, \mu_2 = -2.5, \mu_3 = -1.0\}$, $\{\sigma_1 = 0.3, \sigma_2 = 0.3, \sigma_3 = 0.4\}$.
- CVR sampled from a log-normal distribution $\operatorname{logit} CVR \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \sigma_k^2)$ with $K = 2$, $\{\pi_1 = 0.6, \pi_2 = 0.4\}$, $\{\mu_1 = -2.0, \mu_2 = -1.0\}$, $\{\sigma_1 = 0.7, \sigma_2 = 0.4\}$.
- In the experiment with CTR-only uncertainty, CTR logits were sampled from normal noise with standard deviation from 10^{-2} to 10^{1}.
- In the experiment with joint uncertainty, CVR logits were sampled from normal noise with standard deviation from 10^{-2} to 10^{1}, uncorrelated with noise in CTR logits.
- GMM was fitted via XDGMM on the subsample of 400 auctions in every campaign. The EM algorithm was run for a maximum of 300 iterations or until a log-likelihood tolerance of $2 \cdot 10^{-5}$ was reached.
- For the joint CTR-CVR uncertainty case, we used a 5×5 Gauss-Hermite quadrature grid.

iPinYou

- Number of auctions $T = 1000$, number of campaigns $N = 200$.
- winning price wp sampled from logs.
- CTR sampled from a log-normal distribution $\operatorname{logit} CTR \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \sigma_k^2)$ with $K = 3$, $\{\pi_1 = 0.6, \pi_2 = 0.2, \pi_3 = 0.2\}$, $\{\mu_1 = -3.0, \mu_2 = -2.5, \mu_3 = -1.0\}$, $\{\sigma_1 = 0.3, \sigma_2 = 0.3, \sigma_3 = 0.4\}$.
- CVR sampled from a log-normal distribution $\operatorname{logit} CVR \sim \sum_{k=1}^{K} \pi_k \mathcal{N}(\mu_k, \sigma_k^2)$ with $K = 2$, $\{\pi_1 = 0.6, \pi_2 = 0.4\}$, $\{\mu_1 = -2.0, \mu_2 = -1.0\}$, $\{\sigma_1 = 0.7, \sigma_2 = 0.4\}$.
- In the experiment with CTR-only uncertainty, CTR logits were sampled from normal noise with standard deviation from 10^{-2} to 10^{1}.
- In the experiment with joint uncertainty, CVR logits were sampled from normal noise with standard deviation from 10^{-2} to 10^{1}, uncorrelated with noise in CTR logits.
- GMM was fitted via XDGMM on the subsample of 400 auctions in every campaign. The EM algorithm was run for a maximum of 300 iterations or until a log-likelihood tolerance of $2.0 \cdot 10^{-5}$ was reached.
- For the joint CTR-CVR uncertainty case, we used a 5×5 Gauss-Hermite quadrature grid.

BAT

- Number of campaigns $N = 334$ (and in each campaign individual number of auctions).
- winning price wp sampled from logs.
- CTR sampled from logs.
- CVR sampled from logs.
- In the experiment with CTR-only uncertainty, in CTR logits, normal noise with standard deviation from 10^{-2} to 10^{1} was sampled.
- In the experiment with joint uncertainty, also in CVR logits, normal noise with standard deviation from 10^{-2} to 10^{1} was sampled, uncorrelated with noise in CTR logits.
- GMM was fitted via XDGMM on the subsample of 300 auctions in every campaign. The EM algorithm was run for a maximum of 300 iterations or until a log-likelihood tolerance of $2.0 \cdot 10^{-5}$ was reached.
- For the joint CTR-CVR uncertainty case, we used a 5×5 Gauss-Hermite quadrature grid.

C.2 EXPERIMENTS WITH EMPIRICAL NOISE (CRITEO ATTRIBUTION)

This experiment evaluates the framework's performance using observed model-based uncertainty.

Scenarios:

- **Feature Removal:** We iteratively removed the top-N least important features, increasing predictive uncertainty.
- **Data Scaling:** The models were trained on subsets of the original data (100%, 30%, 10%, 3%, 1%, 0.3%, 0.1%) while maintaining the same validation set for evaluation.

Model Training. We trained `CatBoost` classifiers for both CTR and CVR tasks. We used the Virtual Ensembles mode (`posterior_sampling = True`), which computes the variance of predictions across multiple sub-models in a virtual ensemble.

- **Features:** 12 features (11 of them categorical) including campaign ID, time since last click, and anonymized user metadata.
- **Uncertainty Quantification:** The output variance in logit space $\Sigma_t^2 = \mathrm{diag}({\sigma_t^{\mathrm{CTR}}}^2, {\sigma_t^{\mathrm{CVR}}}^2)$ was used directly as the noise parameter for `DenoiseBid`.

Experiment specification:

- Number of campaigns $N = 505$ (and in each campaign, the individual number of auctions).
- GMM was fitted via XDGMM on the subsample of 400 auctions in every campaign. The EM algorithm was run for a maximum of 300 iterations or until a log-likelihood tolerance of $2.0 \cdot 10^{-5}$ was reached.

D HEAT MAPS IN EXPERIMENTS ON SYNTHETIC AND IPINYOU DATASETS

This section provides visualizations for the joint CTR-CVR uncertainty scenarios on the Synthetic and iPinYou datasets. Figures 6 and 7 illustrate the metrics R/R^* and $\overline{CPC}/\overline{CPC}_{\mathrm{camp}}$ across a range of noise levels in the joint logit space (σ^{CTR} and σ^{CVR}). The observed trends are the same as for the BAT dataset and highlight the performance of `DenoiseBid`.

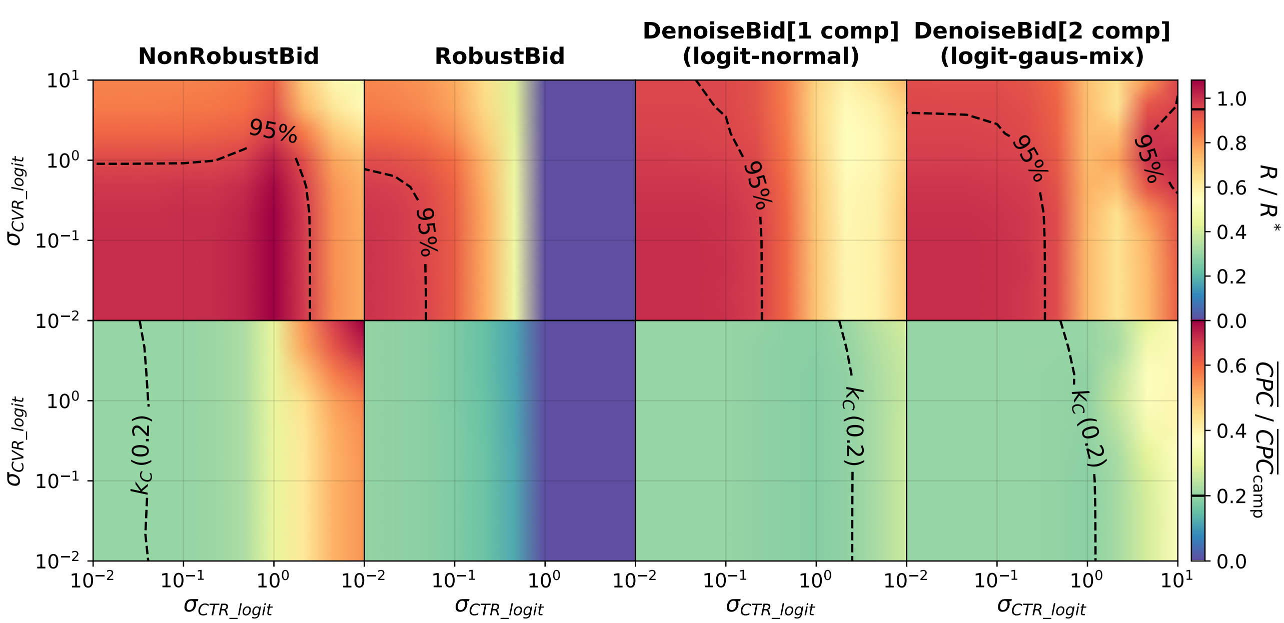

Figure 6: Performance comparison on the synthetic dataset under joint CTR-CVR uncertainty.

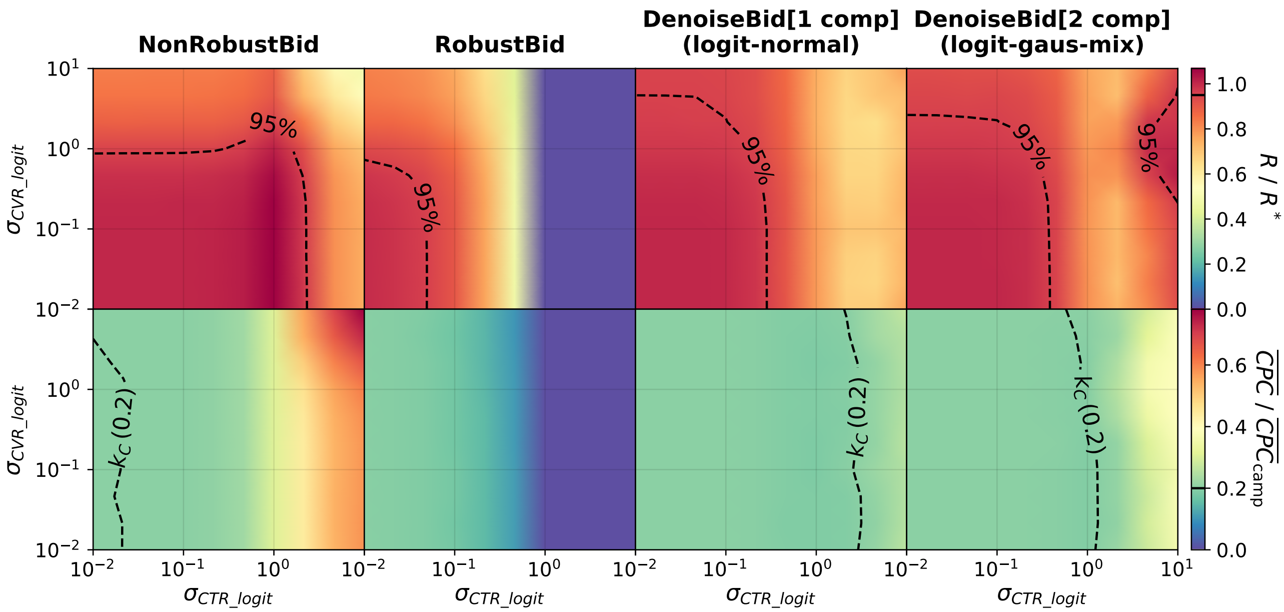

Figure 7: Performance comparison on the iPinYou dataset under joint CTR-CVR uncertainty.

E DEPENDENCE OF THE NUMBER OF GAUSSIANS IN THE GMM PRIOR MODEL

In this section, we investigate how the complexity of the prior distribution, represented by the number of components K in the Gaussian Mixture model, affects the performance of `DenoiseBid`. Figure 8 presents a comparison between the non-robust baseline and `DenoiseBid` with priors ranging from a single Gaussian in logit space (normal prior, $K = 1$) to a 4-component mixture. The experiment was conducted on the BAT dataset.

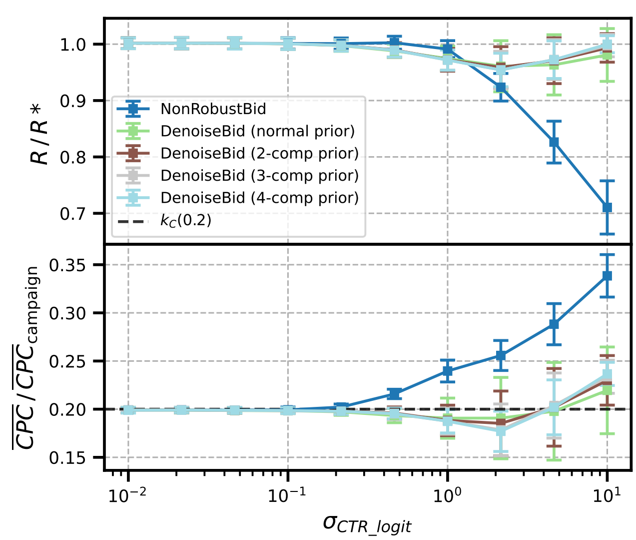

Figure 8: Dependence ofconversion volume and CPC compliance on `DenoiseBid` prior complexity (from 1 to 4 components). The dashed line in the bottom panel represents the target cost-per-click share limit k_C.

While the non-robust baseline fails as noise increases, all denoised variants maintain stability. Multi-component priors ($K \geq 2$) outperform the single Gaussian model ($K = 1$) at high noise regimes ($\sigma_{CTR\_logit} > 1$), however they are slightly worse at adhering to the CPC constraint. Since the performance gap between $K = 2$, $K = 3$ and $K = 4$ is marginal, a mixture with 2 components provides the optimal trade-off between conversions and computational efficiency (see Table 3).