Electronic Companion Supplement

EC.1 Gaussian Process Regression

This section provides an overview of Gaussian process regression. Formally, the assumption is that the demand function D is jointly Gaussian-distributed and completely defined by its mean $\mu(p)$ and covariance function $k(p,p')$, such that $D(p) \sim \text{GP}(\mu(p), k(p,p'))$. The mean and covariance functions are defined as follows (Williams and Rasmussen 2006):

$$\mu(p) = \mathbb{E}[D(p)], \tag{EC.1}$$

$$k(p,p') = \mathbb{E}[(D(p)-\mu(p))(D(p')-\mu(p'))]. \tag{EC.2}$$

For ease of exposition, we set $\mu(p) = 0$.[1]

The kernel can then be used to compute a covariance matrix $K(P,P)$, which contains the covariance between all test points, as well as a covariance matrix (either $K(P,P_t)$ or $K(P_t,P)$) between training and test cases. The joint distribution of the training data P_t and the test points P can be written as follows (equation (2.21) in Williams and Rasmussen (2006)):

$$\begin{pmatrix} y_t \\ D^* \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(P_t,P_t)+\sigma_y^2 I & K(P_t,P) \\ K(P,P_t) & K(P,P) \end{bmatrix}\right), \tag{EC.3}$$

where $D^* = D(P)$ is a random variable denoting the GP predictions at test points P. It then follows from equations (2.22–2.24) in Williams and Rasmussen (2006) that:

$$D^*|P_t, y_t, P \sim N(\mu(D^*), \text{Cov}(D^*)), \text{ where} \tag{EC.4}$$

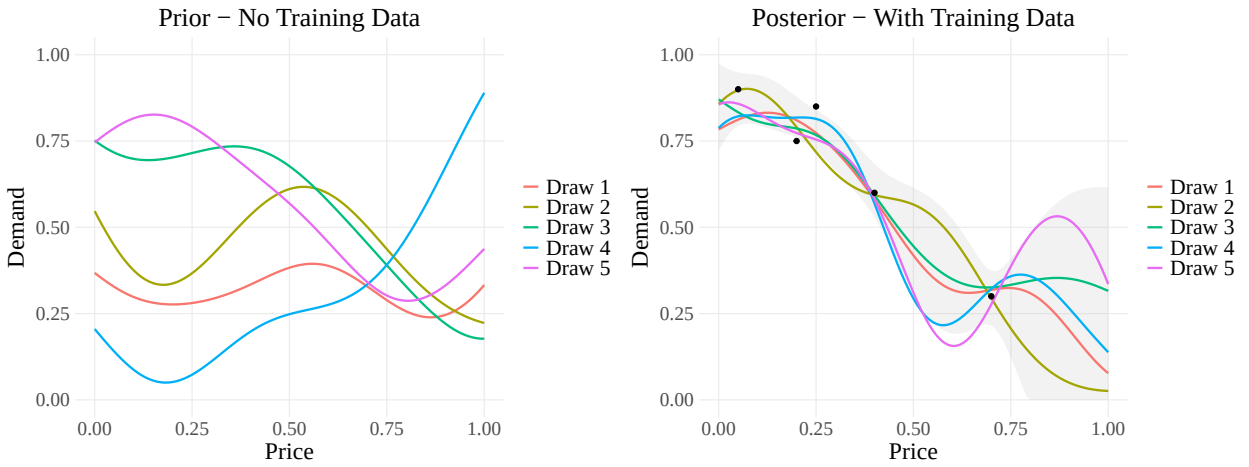
$$\mu(D^*) = K(P,P_t)[K(P_t,P_t)+\sigma_y^2 I]^{-1} y_t, \tag{EC.5}$$

$$\text{Cov}(D^*) = K(P,P) - K(P,P_t)[K(P_t,P_t)+\sigma_y^2 I]^{-1} K(P_t,P). \tag{EC.6}$$

Figure EC.1 illustrates a simple example of GP regression. As data is obtained, the space where the true demand function could exist becomes more constrained. Accordingly, the range of uncertainty is smaller in areas closer to data points compared to those farther away, as shown by the shaded area representing the 95% confidence intervals at the test points.

[1] If we have a GP, $D \sim GP(\mu, k)$, where the prior mean function is non-zero, then $D' = D - \mu$ is the zero-mean Gaussian process, $D' \sim GP(0, k)$. Hence, using observations from the values of D, we can subtract the prior mean function values to obtain observations of D', perform inference on D', and then add back the prior mean $\mu(P)$ to the posterior mean to recover the posterior of D.

Figure EC.1 Random Samples from Gaussian Process With and Without Training Data



Notes. Lines represent five random draws from the GP in both the prior and posterior. In the prior, the mean was set to 0.5. For both the prior and posterior, the RBF kernel was used with hyperparameters $l = 0.2$, $\sigma_D^2 = 0.08$, and $\sigma_y^2 = 0.0016$. There were 101 test points, $P = \{0, 0.01, 0.02, \ldots, 1\}$. The five draws from the posterior distribution were sampled from the GP with training data $P_t = \{0.05, 0.2, 0.25, 0.4, 0.7\}$ and $y_t = \{0.9, 0.75, 0.85, 0.6, 0.3\}$. The shaded area represents the 95% confidence interval at each test point.

EC.2 Computational Issues

A computational issue that arises in fitting a posterior Gaussian process is that matrix inversion is $\mathcal{O}(n^3)$, implying it does not scale well to larger datasets. Typically, the training data grows with each purchase observation (thus depending on t), but we mitigate this issue because purchases can only be observed at prices within the fixed price set. This allows us to set the input training data as the set of test prices and the associated output training data as the observed purchase rates. The only additional adjustment needed is to the noise hyperparameter. As the sample variance scales with the number of observations, the noise hyperparameter is specified accordingly: $\sigma_y^2 = \{0.25/n_{1t}, \ldots, 0.25/n_{At}\}$. This approach ensures that computational complexity does not increase with the number of purchase observations but instead depends on the size of the initial set of test prices.

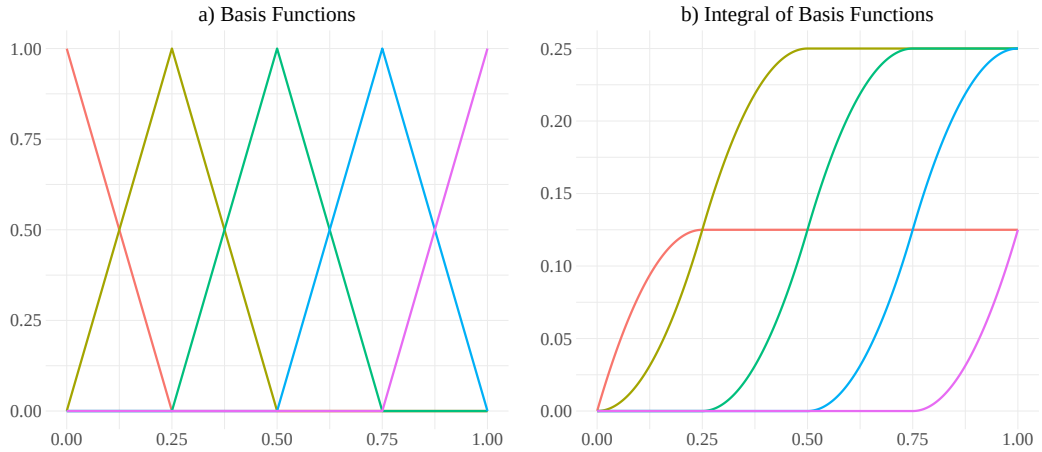
Additionally, one common issue when running GP algorithms is floating-point precision errors, which can lead to negative eigenvalues and violate the positive semi-definite property of covariance matrices. We follow the approach devised by Rebonato and Jäckel (2011) to obtain the nearest covariance matrix.

EC.3 Implementation of Monotonic GP Bandits

EC.3.1 Basis Function Visualization

Consider an example where $N = 4$ meaning there are 5 equally spaced knots $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$. Figure EC.2 shows the five basis functions along with their corresponding integrals.

Figure EC.2 Plot of Basis Functions and their Integrals (5 knots)



EC.3.2 Derivation of Proposition 1

LEMMA EC.1. *The distance between a continuous function D and its estimate via basis functions $D_J(p) = \sum_{j=0}^{J} D(u_j)h_j(p)$ converges to 0 as $J \to \infty$.*

From Lemma EC.1, a continuous demand function D can be estimated via basis functions as $D(p) \approx \sum_{j=0}^{J} D(u_j)h_j(p)$. Since, by assumption, the derivative of D is also continuous, the same formula can be used to estimate D' as follows:

$$D'(p) \approx \sum_{j=0}^{J} D'(u_j)h_j(p). \tag{EC.7}$$

Additionally, by the Fundamental Theorem of Calculus,

$$D(p) - D(0) = \int_0^p D'(t)\,dt. \tag{EC.8}$$

Substituting (EC.7) into (EC.8) gives

$$D(p) \approx D(0) + \sum_{j=0}^{J} D'(u_j) \int_0^p h_j(t)\,dt. \quad \square \tag{EC.9}$$

EC.3.3 Gaussian Process – Estimation of Derivatives and Intercept

The basis function method requires the estimation of the intercept and the derivatives at each of the prices in the consideration set. More formally, the goal is to estimate the posterior mean and covariance for $\{D(0), D'(p_1), \ldots, D'(p_A)\}$. Temporarily ignoring the intercept, the key insight is that the derivatives of a GP are also a GP. This implies that D'^*, the posterior vector of derivatives of D^*, is given by:

$$D'^* \sim N\left(\frac{d}{dp}\mu(D^*), \frac{d}{dp}\mathrm{Cov}(D^*)\right). \tag{EC.10}$$

Since the values of D^* are only at the test points P, the derivative only needs to be calculated with respect to P. Consequently, to estimate the posterior mean and covariance for $\{D(0), D'(p_1), \ldots, D'(p_A)\}$, the only necessary adjustment is to compute the derivatives of the kernel function with respect to the test points.

We compute the partial derivatives of the kernel with respect to the prices as follows:

$$\frac{\partial k(p_i^*, p_j)}{\partial p_i^*} = \frac{\sigma_D^2}{l^2}(p_j - p_i^*)\exp\left(\frac{-(p_i^* - p_j)^2}{2l^2}\right), \tag{EC.11}$$

$$\frac{\partial k(p_i, p_j^*)}{\partial p_j^*} = \frac{\sigma_D^2}{l^2}\left(p_i - p_j^*\right)\exp\left(\frac{-(p_i - p_j^*)^2}{2l^2}\right), \tag{EC.12}$$

$$\frac{\partial^2 k(p_i^*, p_j^*)}{\partial p_i^* \partial p_j^*} = \frac{\sigma_D^2}{l^4}\left(l^2 - (p_i^* - p_j^*)^2\right)\exp\left(\frac{-(p_i^* - p_j^*)^2}{2l^2}\right). \tag{EC.13}$$

EC.4 Regret Bounds

Consider a setting where $P = \{p_1 \le p_2 \le \cdots \le p_d\}$ are a subset of prices that we choose as knots. We consider a Gaussian process for which draws are C^1 almost surely, and consider the joint distribution over draws f, f'. We define the set of monotonic functions,

$$\mathcal{M} = \{f \in C^1([0,1]) : f'(x) \le 0, x \in [0,1]\}.$$

Ideally we would like to restrict draws of our GP to $\mathcal{M}$, but in general, since we can only evaluate our GP at a finite set of points, we instead insist that our function is monotonic at the set of knots,

$$\mathcal{M}(P) = \{f(\in C^1[0,1] : f'(p) \le 0, p \in P\}$$

We denote the joint prior distribution over the function and the derivative $\Pi_0 = GP([0,0], K|\mathcal{M}(P))$, where K is the appropriate kernel.

Let $p^* = \arg\max_{p \in P} f(p)$ – note that since f is drawn from the underlying prior, p^* is a random variable. We define the Bayesian regret of our policy

$$BR_T = \sum_{t=1}^{T} \mathbb{E}[p^* f(p^*) - p_t f(p_t)]$$

where the expectation is over draws from the prior, reward noise, and any internal randomness of the algorithm.

Throughout the following, we refer to the truncated distribution as Π_t, and the untruncated distribution $GP([\mu_t, \mu_t'], K_t)$ as $\Pi_{t,u}$. Let the induced probability laws and expecation, with respect to the measure conditioned on the history $\mathcal{H}_t = \{(p_s, r_s)\}_{s=1}^t$, be $\mathbb{P}_t, \mathbb{E}_t$, and for the untruncated version, $\mathbb{P}_{t,u}, \mathbb{E}_{t,u}$. Critically we make the following assumption:

Assumption 1. The probability of returning a function monotonic on the knots is bounded below, i.e., there exists $c \geq 0$ such that

$$\mathbb{P}_{t,u}(f_t, f_t' \in \mathcal{M}(P)) \geq c, \forall t \geq 1$$

Remark: Note that $\mathbb{P}_{t,u}(f_t'(P) \geq 0)$ is equivalent to $\mathbb{P}(y \geq 0)$ for $x, y \sim N([\mu_t(P), \mu_t'(P)], K_t)$. This is an integral of a multivariate Gaussian over an open set. Since $f'(P) \geq 0$ by definition of the prior, if $\mu'(P) \to f'(P)$, we should expect this probability to actually increase with t.

Additionally, we require the following lemma linking a distribution with its truncated version.

LEMMA EC.2. *Let $p(x)$ be a distribution on $\mathbb{R}^d$. Let $S \subset \mathbb{R}^d$. Define the truncated pdf on $\mathbb{R}^d$, $p_S(x) = \mathbf{1}\{x \in S\}p(x)/\mu(S)$ where $\mu(S) = \int_{x \in S} p(x)$. Then given an event $E \subset \mathbb{R}^d$,*

$$\mathbb{P}_p(\mathbf{1}\{E \cap S\}) \leq \mathbb{P}_{p_S}(E) \leq \frac{1}{\mu(S)}\mathbb{P}_p(E)$$

and given a function $f(x) : \mathbb{R}^d \to \mathbb{R}$

$$\mathbb{E}_{p_S}(f(x)) \leq \frac{1}{\mu(S)}\mathbb{E}_p(f(x))$$

Proof. The lower bound is immediate since $\mu(S) \leq 1$.

$$\begin{aligned}
\mathbb{P}_{p_S}(E) &= \int_{x \in \mathbb{R}^d} \mathbf{1}\{x \in E\}\mathbf{1}\{x \in S\}p(x)/\mu(S) \\
&= \frac{1}{\mu(S)} \int_{x \in \mathbb{R}^d} \mathbf{1}\{x \in E\}\mathbf{1}\{x \in S\}p(x) \\
&\leq \frac{1}{\mu(S)} \int_{x \in \mathbb{R}^d} \mathbf{1}\{x \in E\}p(x) \\
&\leq \frac{1}{\mu(S)}\mathbb{P}_p(E)
\end{aligned}$$

The result on expectations is immediate. □

THEOREM EC.1. *The Bayes Regret of Algorithm 1 GP-TS-M is bounded by*

$$\begin{aligned}
BR_t &\leq \mathbb{E}\left[\sqrt{\sum_{t=1}^{\infty} p_t^2}\right] \sqrt{\gamma_T \log(1 + \sigma_0^{-2}) \log\left(\frac{T^2|A|}{\sqrt{2\pi}}\right)} + \mathcal{E}_T + 1 \\
&\leq \sqrt{T\gamma_T \log(1 + \sigma_0^{-2}) \log\left(\frac{T^2|A|}{\sqrt{2\pi}}\right)} + \mathcal{E}_t + 1
\end{aligned}$$

where $\mathcal{E}_T = \sum_{t=1} \mathbb{E}[\mu_t(p^*) - \mathbb{E}_t[f_t(p^*)]]$ and γ_T is the mutual information of the Gaussian process (Srinivas et al. 2009).

Interpreting the Regret: We remark that if we were not restricting to the monotonic set of functions, the argument shows that $\mathcal{E}_T$ is 0. And so the final regret is of the form $O(\sqrt{\gamma_T T \log(T|P|)})$. Note that this regret is independent of the underlying constraint set of monotonic functions. To understand the impact of the underlying constraint set, we focus on the path-dependent regret term $\sum_{t=1}^{T} p_t^2$. In general, this quantity is less than the maximum price played times T, and is a tighter regret result compared to existing works. We remark that the looseness in this result is primarily due to using loose tail bounds that do not effectively account for the constraint set. Future work could examine different and potentially tighter bounds.

Discussion of $\mathcal{E}_T$. In Figure EC.3, we have plotted $\log(t)$ vs. $\log(E_t)$, where E_t estimates each term of $\mathcal{E}_t$, defined as $\mathbb{E}[\mu_t(p^*) - \mathbb{E}_t[f_t(p^*)]]$. At each update of the Gaussian process, f_t is obtained from 10,000 posterior draws, and the difference from μ_t is calculated to estimate E_t. To ensure robustness, these E_t values were averaged across 1,000 independent simulations. As the plot demonstrates, $E_t \approx t^{-\alpha}$ where $\alpha \in [.5, .8]$. This implies that $\mathcal{E}_T \leq O(T^{-\alpha+1})$ hence contributing a regret bounded by $O(\sqrt{T})$. Intuitively this is not surprising, we should expect fast concentration of $\mu_t(p^*)$ and $\mathbb{E}_t[f_t(p^*)]$ to both quickly concentrate to $\mu(p^*)$ at a rate that matches the parametric rate of $1/\sqrt{t}$.

Proof. Define $U_t(p) := \mu_{t-1}(p) + \beta_{t-1}^{1/2}\sigma_{t-1}(p)$ where $\beta_t = \log(t^2 c^{-1}|P|/\sqrt{2\pi})$. Note that, conditioned on $\mathcal{H}_t$, the optimal action p^* and the action p_t selected by posterior sampling are identically distributed by Fact 5 (see below). In addition, U_t is deterministic conditioned on the history, so, $\mathbb{E}_t[U_t(p^*)] = \mathbb{E}_t[U_t(A_t)]$. Therefore,

$$\begin{aligned}
\mathbb{E}[p^* f(p^*) - p_t f(p_t)] &= \mathbb{E}[\mathbb{E}_t[p^* f(p^*) - p_t f(p_t)]] \\
&= \mathbb{E}[\mathbb{E}_t[p_t U_t(p_t) - p^* U_t(p^*) + p^* f(p^*) - p_t f(p_t)]] \\
&= \mathbb{E}[\mathbb{E}_t[p_t U_t(p_t) - p_t f(p_t)] + \mathbb{E}_t[p^* f(p^*) - p^* U_t(p^*)]] \\
&= \mathbb{E}[p_t U_t(p_t) - p_t f(p_t)] + \mathbb{E}[p^* f(p^*) - p^* U_t(p^*)].
\end{aligned}$$

Thus, we see that we can bound the Bayes-Regret as

$$BR(T) \leq \sum_{t=1}^{T} \mathbb{E}[p_t U_t(p_t) - p_t f(p_t)] + \sum_{t=1}^{T} \mathbb{E}[p^* f(p^*) - p^* U_t(p^*)] \tag{EC.14}$$

(EC.15)

We now focus on the first term,

$$\begin{aligned} p_t U_t(p_t) - p_t f(p_t) &= p_t U_t(p_t) - p_t \mu_t(p_t) + p_t \mu_t(p_t) - p_t f(p_t) \\ &= \mathbb{E}[p_t U_t(p_t) - p_t \mu_t(p_t)] + p_t \mu_t(p_t) - p_t f(p_t) \\ &\leq p_t \beta_t^{1/2} \sigma_t(p_t) + p_t \mu_t(p_t) - p_t f(p_t) \\ &\leq p_t \beta_t^{1/2} \sigma_t(p_t) + \mu_t(p_t) - f(p_t) \end{aligned}$$

Next,

$$\sum_{t=1}^{T} p_t \beta_t^{1/2} \sigma_t(p_t) \leq \sqrt{\beta_T \sum_{t=1}^{\infty} p_t^2} \sqrt{\sum_{t=1}^{\infty} \sigma_t^2(p_t)} \qquad \text{(Cauchy-Schwartz)}$$

A standard argument (see Srinivas et al. (2009)) shows that

$$\sum_{t=1}^{\infty} \sigma_t^2(p_t) \leq \frac{\gamma_T}{\log(1+\sigma^{-2})}$$

Finally, we bound the second term of EC.14

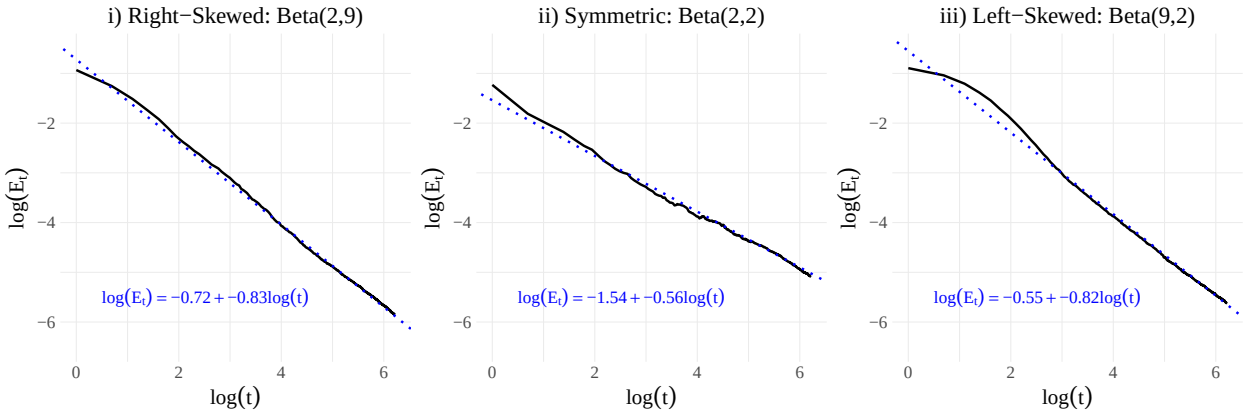
$$\begin{aligned} \sum_{t=1}^{T} E[p^* f(p^*) - p^* U_t(p^*)] &\leq \sum_{t=1}^{\infty} \sum_{p \in P} \mathbb{E}_t[\mathbf{1}\{f(p) - U_t(p) \geq 0\}(f(p) - U_t(p))] \\ &\leq \sum_{t=1}^{\infty} \sum_{p \in P} \frac{1}{\mathbb{P}_{t,u}(M)} \mathbb{E}_{t,u}[\mathbf{1}\{f(p) - U_t(p) \geq 0\}(f(p) - U_t(p))] \\ &\leq \sum_{t=1}^{\infty} \sum_{p \in P} \frac{1}{c} \mathbb{E}_{t,u}[\mathbf{1}\{f(p) - U_t(p) \geq 0\}(f(p) - U_t(p))] \end{aligned}$$

Now, in the untruncated distribution, $\mathbb{E}_{t,u}[f(p) - U_t(p)] = -\beta_t^{1/2} \sigma_t^2(p)$, which is negative. Thus, using standard tail bounds Russo and Van Roy (2014),

$$\begin{aligned} \sum_{t=1}^{T} E[p^* f(p^*) - p^* U_t(p^*)] &\leq \frac{1}{c} \sum_{t=1}^{\infty} \frac{\sigma_t(p)}{\sqrt{2\pi}} e^{-\beta/2} \\ &\leq \frac{1}{c} \sum_{t=1}^{\infty} \frac{\sigma_t(p)}{t^2 |P| c^{-1}} \leq 1 \end{aligned}$$

The result follows from combining all the terms.

Figure EC.3 Decay of E_t under Different Distributions



Notes. At each update step of the Gaussian process, E_t is calculated as the expected difference $\mathbb{E}[\mu_t(p^*) - \mathbb{E}_t[f_t(p^*)]]$, where f_t is obtained averaging over 10,000 posterior draws. Results are averaged over 1,000 independent simulations. The solid black lines show the empirical estimates of $\log(E_t)$, while the blue dotted lines display the corresponding linear fits, with equations annotated in each panel.

EC.5 Simulations Using Alternative WTP Distributions

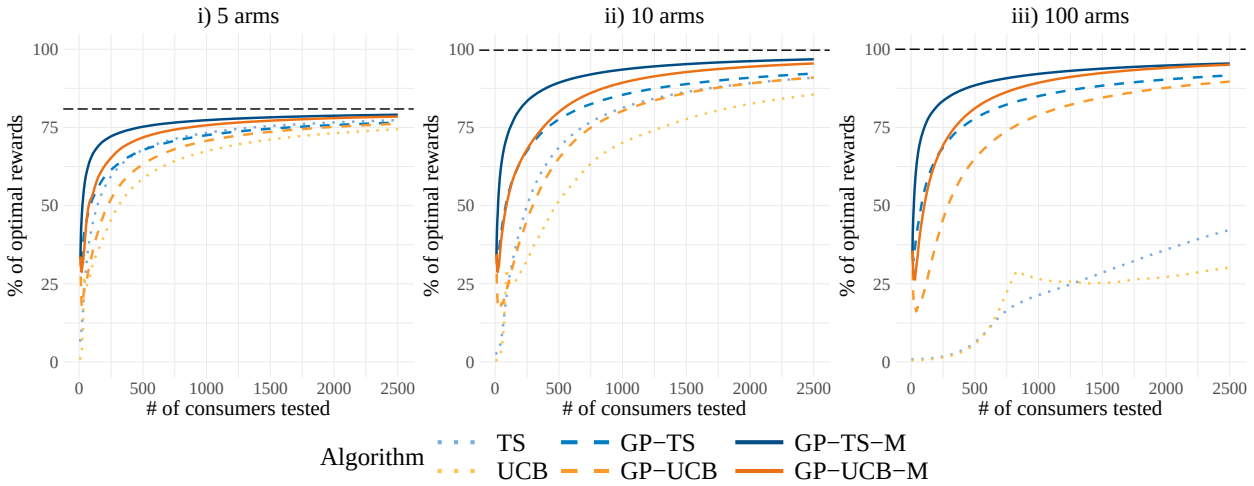
We evaluate the robustness of the proposed method across a variety of challenging scenarios to demonstrate its practical value.

EC.5.1 Field Data

To further demonstrate the applicability of our method, we tested it on real-world data. The data comes from an empirical study of demand for a music streaming subscription service where the distribution of WTP for a monthly plan is estimated (Chou and Kumar 2024; see Figure 2). From this, we then normalize the WTP distribution to lie in the range $[0, 1]$; specifically, $p' = \frac{p}{1000}$. Using this WTP distribution, we run the bandit algorithms using the same setup as in the simulations.

Since the optimal price is relatively low (0.21) within the price set, we expect similar uplifts from the informational externalities as observed in the Beta(2,9) case from the main analysis. For the first informational externality, the uplift increases with the number of arms, being slightly negative for 5 arms, slightly positive for 10 arms, and dramatically positive for 100 arms. For the second informational externality, across all price sets the uplifts are positive and similar in size. Additionally, GP-TS-M is the best performing algorithm. These findings are consistent with the Beta(2,9) simulations and highlight the validity and practical value of our method.

Figure EC.4 Field Data: Cumulative Percent of Optimal Rewards (Profits)



Notes. The lines represent the means of the cumulative expected percentage of optimal rewards across 1000 simulations. The black horizontal line represents the maximum obtainable reward given the price set, while 100% represents the true optimal reward given the underlying distribution.

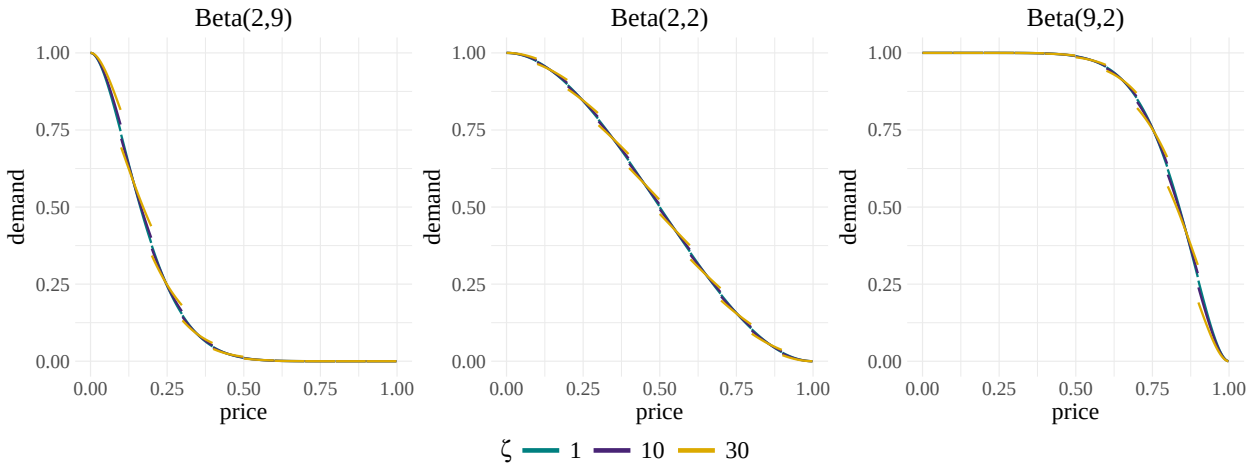
EC.5.2 Discontinuous Demand: Left-Digit Bias

We next discuss the case where consumers may be affected by left-digit bias. This phenomenon occurs when the demand function is discontinuous at a price where the left digit changes (e.g., from \$1.99 to \$2.00), as consumers perceive the price increase to be larger than one cent (Thomas and Morwitz 2005). For instance, Strulov-Shlain (2023) empirically found that consumers reacted to a one-cent increase from a ninety-nine-ending price (resulting in a left-digit change) as if it were a twenty-cent increase. This is an important case to test, as GP-based models rely on continuity and may struggle with discontinuities.

To be consistent with left-digit literature, we discretize the continuum of prices $[0, 1]$ into 1000 points and introduce discontinuities at changes in the left-most significant digit (e.g., 0.099 to 0.100, 0.199 to 0.200, etc.). This ensures the left-digit effect occurs between prices ending in ninety-nine and zero. To specify the size of the discontinuities, we calculate the difference in demand between consecutive prices where the left-digit changes and multiply it by a scale factor ζ. For instance, if $\zeta = 20$, the gap is 20 times larger than usual – consistent with Strulov-Shlain (2023). We then rescale the continuous portion of the demand curve to accommodate these gaps.

For our simulations, we used demand curves (Figure EC.5) derived from three WTP distributions – Beta(2,9), Beta(2,2), and Beta(9,2) – adjusted with $\zeta = 10$ (low left-digit bias) and $\zeta = 30$ (high left-digit bias). These cases provide generous bounds for estimates of left-digit bias based on Strulov-Shlain (2023).

Figure EC.5 Left-Digit Demand Curves



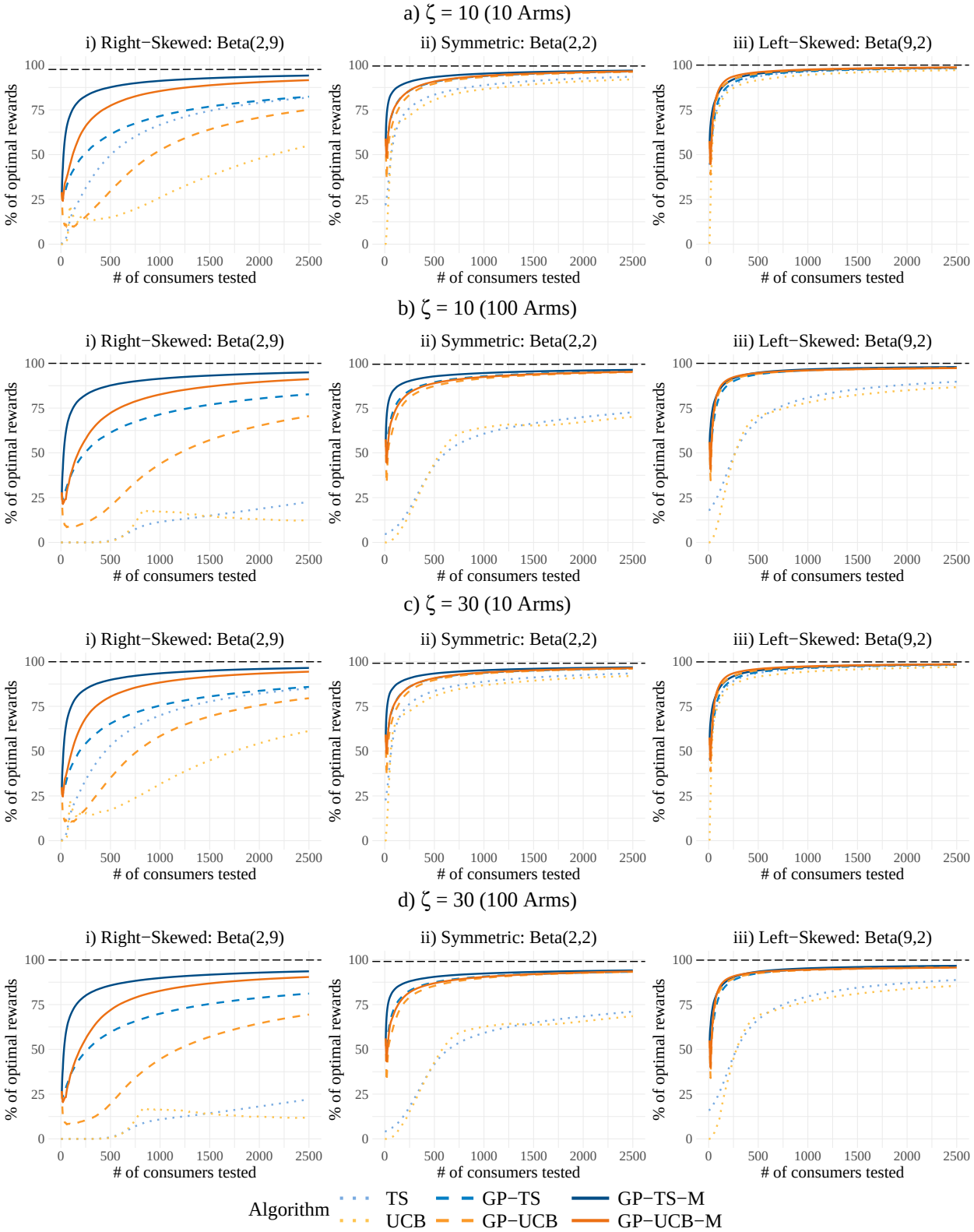
Notes. Price is discretized at intervals of 0.001 from 0 to 1, with left-digit discontinuities occurring every 0.01. $\zeta = 1$ represents the base case with no left-digit bias (discontinuities arise solely from price discretization). $\zeta = 10$ corresponds to low left-digit bias, with a discontinuity gap 10 times larger than the base case. $\zeta = 30$ corresponds to high left-digit bias, with a discontinuity gap 30 times larger than the base case.

Our simulation results are shown in Figure EC.6. With 10 arms, we observe that in both cases (low ζ and high ζ), GP-based algorithms perform as well as in the main analysis. That is with just 10 arms, left-digit discontinuities do not affect GP-based algorithms. Perhaps surprisingly, this outcome is expected as when only 10 prices are tested, each price falls within a distinct interval of the piecewise function, leaving enough distance between prices for the discontinuities to have no impact.

The issue arises when multiple prices are tested within each interval, requiring the GP to learn both the continuous segments and the discontinuity gaps. Since the GP assumes continuity, it tends to smooth over these gaps, resulting in a slight misestimation of the demand curve. For example, in the Beta(2,2) case, when there are low left-digit discontinuity gaps ($\zeta = 10$), GP-TS-M achieves 96.9% (after 2500 consumers) of the true optimal for 10 arms and 96.3% for 100 arms – a slight decrease. However, when there are high left-digit discontinuity gaps ($\zeta = 30$), this decline is more pronounced (96.8% to 94.1%). This is because as the GP tries to smooth over larger discontinuity gaps, it leads to greater misestimation and the selection of slightly suboptimal prices. This can be seen in subfigure d)ii) of Figure EC.6, where the cumulative optimal rewards curve for GP-TS-M flattens before reaching the optimal, as the algorithm prefers to select 0.43, which provides 3% less reward than the true optimal price of 0.39.

Despite the GP's difficulty in handling discontinuities with 100 arms, the monotonic versions of the algorithm remain the best performers after 2500 consumers. While TS and UCB

Figure EC.6 Left Digit: Cumulative Percent of Optimal Rewards (Profits)



Notes. The lines represent the means of the cumulative expected percentage of optimal rewards across 1000 simulations. The black horizontal line represents the maximum obtainable reward given the price set, while 100% represents the true optimal reward given the underlying distribution. ζ is a measure of the size of the discontinuity gap that occurs at locations where the left-digit changes.

(which can handle discontinuities as they model each arm independently) will eventually learn the optimal price and surpass the performance of GP-TS-M and GP-UCB-M, this will only happen for extremely large customer counts. For any reasonable number of tested consumers, the additional exploration cost from forgoing the informational externalities far outweighs the small performance loss from slight misestimation.

To summarize, even with left-digit bias, our algorithms continue to perform best empirically. From a managerial perspective, if a firm suspects left-digit bias, a practical approach is to test prices only at the discontinuities (e.g., 1.99, 2.99, etc.) to avoid misspecification issues with the GP. However, testing fewer prices may lower the maximum reward obtainable from the price set. Thus, the trade-off in determining the number of prices to test involves balancing the potential for a higher possible maximum reward against the risk of misestimation. This trade-off is evident in our experiments: in the Beta(2,9) case, when 100 prices were tested instead of 10, GP-TS-M performed better when left-digit bias was low [subfigure a)i) vs. b)i)] but worse when left-digit bias was high [subfigure c)i) vs. d)i)].

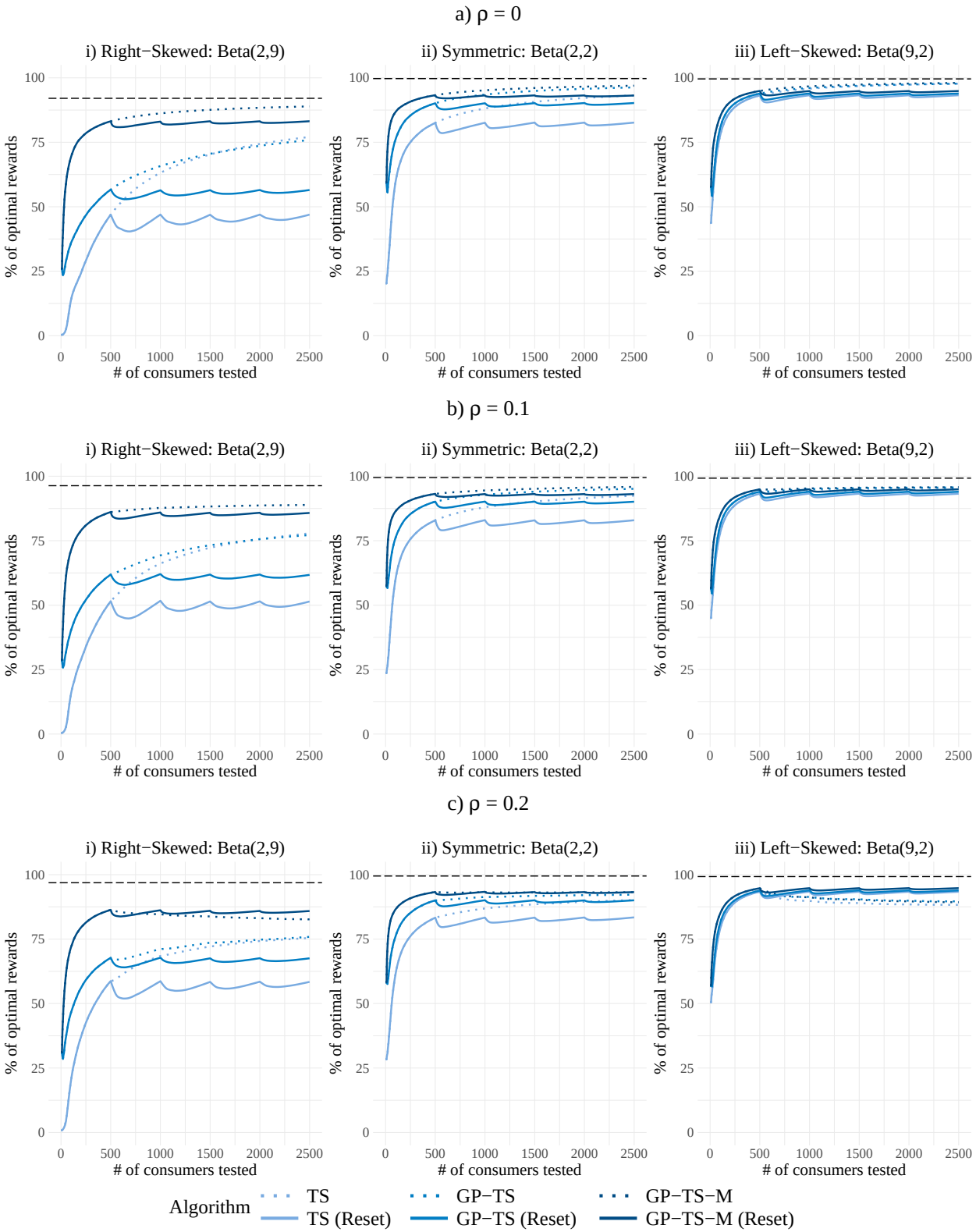
EC.5.3 Time Varying Demand

We now consider the case where demand changes depending on the season (Soysal and Krishnamurthi 2012). We model this by introducing a shock (drawn from the uniform distribution $[-\rho, \rho]$) to the underlying WTP distribution every Q consumers. Specifically, the WTP distribution is shifted horizontally by the value of the shock, increasing or decreasing every consumer's valuation by this amount and resulting in a horizontal shift in the demand curve.

We evaluate the performance (results are shown in Figure EC.7) of the proposed algorithms under these time-varying distributions in two ways. First, we consider the usual baselines (represented by dotted lines), which retain all data from the experiment. Alternatively, we analyze reset variants (represented by solid lines), where the learning process is reset every 500 consumers by discarding all previous experiment data.

First, we analyze how performance changes as ρ varies. When there are no seasonal shocks ($\rho = 0$), resetting results in worse performance for all algorithms. However, the performance gap between the baseline and its corresponding reset variant decreases as informational externalities are incorporated. Furthermore, as ρ increases, the performance decrease from using the reset variants diminish, with GP-TS-M reset variants actually outperforming the baselines when ρ is sufficiently large ($\rho = 0.2$).

Figure EC.7 Time Varying: Cumulative Percent of Optimal Rewards (10 arms)



Notes. The lines represent the means of the cumulative expected percentage of optimal rewards across 1000 simulations. A demand shock is drawn from $[-\rho, \rho]$ every 500 consumers. The black horizontal line represents the maximum obtainable reward given the price set, while 100% represents the true optimal reward given the underlying distribution.

Next, we analyze how performance differs across distributions. The performance decrease from using the reset variants instead of the baselines are most pronounced for Beta(2,9) and decrease as the distribution shifts to Beta(2,2) and then to Beta(9,2). In the Beta(9,2) case, when demand shocks are sufficiently large ($\rho = 0.2$), the reset variants outperform the baselines for all three algorithms. However, for the same ρ under Beta(2,9), the TS and GP-TS reset variants perform significantly worse than the baselines, with only GP-TS-M performing better with a reset.

These patterns can be understood by considering the trade-off of resetting. Resetting incurs a cost in learning, as the algorithm must re-learn the demand from scratch. Conversely, resetting enables more accurate learning when the underlying demand has shifted. Thus, resetting performs better only when the losses from learning a new demand curve are less than the losses from using pre-shock data. This suggests that resetting is most applicable when demand shocks are large (i.e., ρ is large) and when learning is relatively easy (i.e., when the optimal price is high within the price set, as in Beta(9,2)).

Finally, regardless of the size of the shocks or the underlying demand curve, resetting is consistently more effective for algorithms with higher learning efficiency. Since incorporating informational externalities enhances learning efficiency, these externalities provide a persistent advantage to the algorithms that leverage them.

EC.6 Heteroscedastic Noise

As discussed in the main results, when there are very few arms, TS can outperform GP-TS. This is intuitive, as the value of the first informational externality (learning across arms) decreases when there are fewer arms that are further apart. Another key consideration is that TS learns the noise separately for each arm, whereas standard GPs assume homoscedastic noise across all arms.

This assumption is a limitation of standard GPs, as the noise around purchase rates is inherently heteroscedastic. Intuitively, the sample mean at a price where nearly every consumer either purchases or does not purchase is much less noisy than at a price where consumers are equally likely to purchase and not purchase. Specifically, since purchase is a binary decision, the variance at purchase rate y_{at} is given by $y_{at}(1 - y_{at})$. This appendix addresses this limitation by introducing an alternative method for tuning the noise hyperparameter to account for heteroscedasticity.

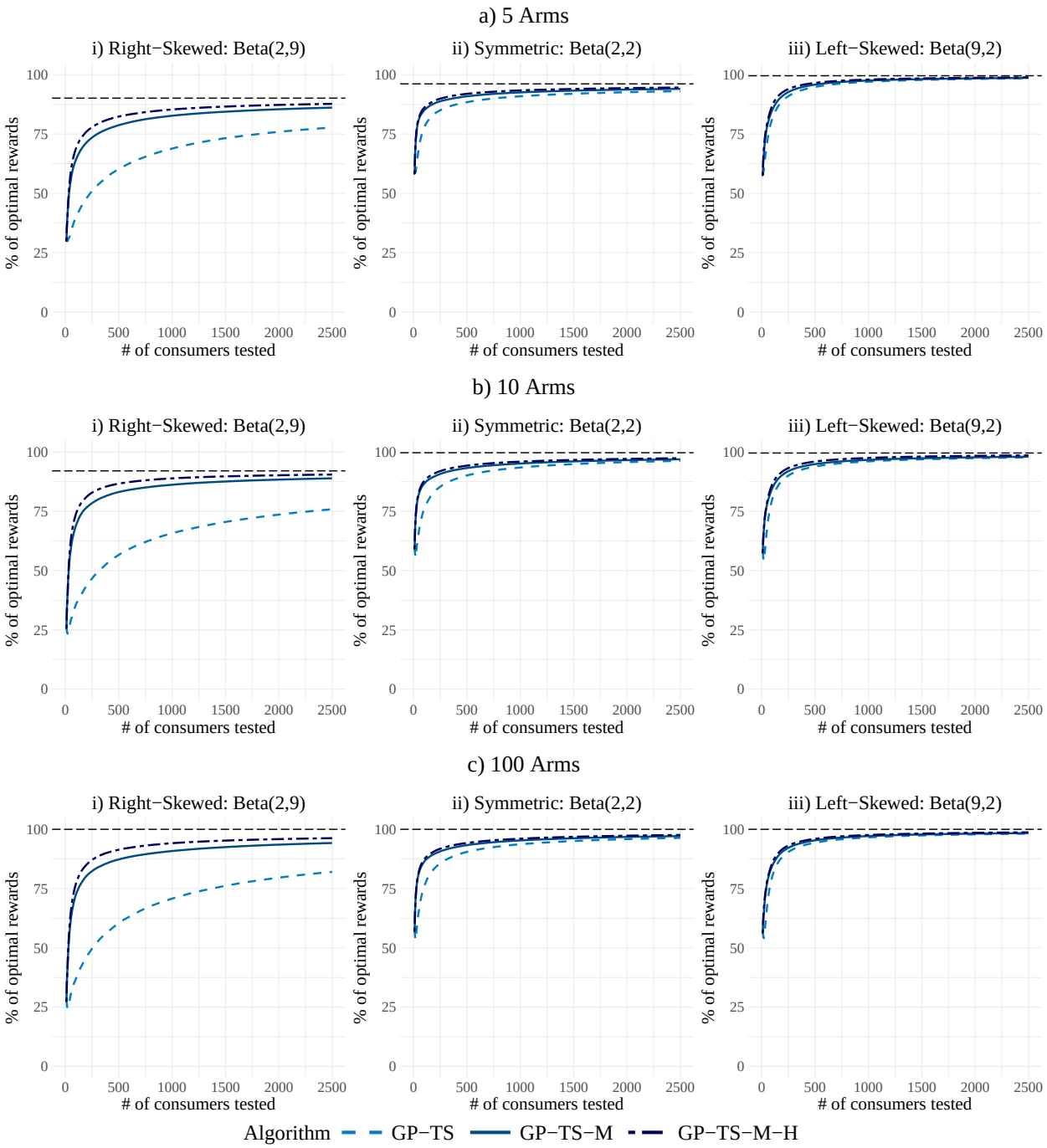
Implementation: While there are methods to estimate a GP with heteroscedastic noise (Goldberg et al. 1997, Kersting et al. 2007), they face the same challenges as estimating noise using MLE under the homoscedastic specification. Specifically, it can be difficult to accurately identify the shape and noise hyperparameters (Murray 2008), which poses significant problems for bandits, as an insufficient noise estimate can cause the algorithm to get stuck on suboptimal arms.

An alternative approach is to model the error by specifying an underlying structural process. In our case, the noise can be modeled based on how likely a consumer is to purchase. Generally, this can be written as $\sigma_y^2 = g(D(p))$ for some unknown function g. However, because the noise depends on the underlying distribution, which is completely unknown, it is unlikely that any suitable candidates for $g(\cdot)$ exist.

Our estimation process for heteroscedastic noise operates as follows. First, we estimate the GP using homoscedastic noise, producing the standard results. Next, we take a demand draw from this GP, which (due to the binary nature of purchase decisions) allows the noise estimate at each price to be calculated as $\tilde{D}(p)(1-\tilde{D}(p))$. Using this noise input, we estimate another GP, after which the bandit process continues as usual. Importantly, as this method updates with a new noise draw at each update iteration, it provides more accurate noise estimates while still preventing the algorithm from getting stuck from consistently underestimating noise. We refer to this implementation as *GP-TS-H*, with its monotonic version called *GP-TS-M-H*.

Results: The results are presented in Figure EC.8. Including heteroscedasticity in addition to monotonicity results in a small performance increase across all simulations. As with other informational externalities, the effects are most pronounced in the Beta(2,9) case. This is because smaller noise hyperparameters synergize with monotonicity to reduce the space of potential demand curves in the low-reward, high-price region.

Figure EC.8 Heteroscedasticity: Cumulative Percent of Optimal Rewards (Profits)



Notes. The lines represent the means of the cumulative expected percentage of optimal rewards across 1000 simulations. The black horizontal line represents the maximum obtainable reward given the price set, while 100% represents the true optimal reward given the underlying distribution.

References

Chou C, Kumar V (2024) Estimating demand for subscription products: Identification of willingness to pay without price variation. *Marketing Science* .

Goldberg PW, Williams CK, Bishop CM (1997) Regression with input-dependent noise: A gaussian process treatment. *Advances in neural information processing systems* 10:493–499.

Kersting K, Plagemann C, Pfaff P, Burgard W (2007) Most likely heteroscedastic gaussian process regression. *Proceedings of the 24th international conference on Machine learning*, 393–400.

Murray I (2008) Introduction to gaussian processes. *Dept. Computer Science, University of Toronto* .

Rebonato R, Jäckel P (2011) The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. *Available at SSRN 1969689* .

Russo D, Van Roy B (2014) Learning to optimize via posterior sampling. *Mathematics of Operations Research* 39(4):1221–1243.

Soysal GP, Krishnamurthi L (2012) Demand dynamics in the seasonal goods industry: An empirical analysis. *Marketing Science* 31(2):293–316.

Srinivas N, Krause A, Kakade SM, Seeger M (2009) Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995* .

Strulov-Shlain A (2023) More than a penny's worth: Left-digit bias and firm pricing. *Review of Economic Studies* 90(5):2612–2645.

Thomas M, Morwitz V (2005) Penny wise and pound foolish: the left-digit effect in price cognition. *Journal of Consumer Research* 32(1):54–64.

Williams CK, Rasmussen CE (2006) *Gaussian processes for machine learning*, volume 2 (MIT press Cambridge, MA).