Supplementary Material

S-1. Proof of Lemma 1

We first derive an upper bound for the expected cumulative regret. We use $\pi(t)$ to denote the arm that is pulled in period t and $T_{i,j}$ is the number of times that arm i is pulled during first j periods. The time horizon n will be fixed in the following proof. Given the definition of the empirical mean-variance in (3), we rewrite the empirical mean as follows,

$$\widehat{\mu}_n(\pi) = \frac{1}{n}\sum_{t=1}^{n} X_{\pi(t),t} = \frac{1}{n}\sum_{i=1}^{K} T_{i,n}\widehat{\mu}_{i,T_{i,n}}, \qquad \mu_1 = \frac{1}{n}\sum_{i=1}^{K} T_{i,n}\mu_1$$

where $\widehat{\mu}_{i,T_{i,n}} = \frac{1}{T_{i,n}}\sum_{t=1}^{n} X_{\pi(t),t}\mathbf{1}_{\pi(t)=i}$.

Similarly, the variance term can be written as

$$\widehat{\sigma}_n^2(\pi) = \frac{1}{n}\sum_{t=1}^{n}\left(X_{\pi(t),t} - \widehat{\mu}_n(\pi)\right)^2 = \frac{1}{n}\sum_{i=1}^{K} T_{i,n}\widehat{\sigma}^2_{i,T_{i,n}} + \frac{1}{n}\sum_{i=1}^{K} T_{i,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_n(\pi)\right)^2.$$

We can further bound the second term as follows,

$$\frac{1}{n}\sum_{i=1}^{K} T_{i,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_n(\pi)\right)^2 \leq \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i} T_{i,n}T_{j,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_{j,T_{j,n}}\right)^2.$$

Then

$$\mathcal{R}_n(\pi) \leq \sum_{i=2}^{K} T_{i,n}\left(\mathrm{MV}_1 - \widehat{\mathrm{MV}}_{i,T_{i,n}}\right) + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i} T_{i,n}T_{j,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_{j,T_{j,n}}\right)^2.$$

Taking the expectation of the right hand side, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i=2}^{K} T_{i,n}\left(\mathrm{MV}_1 - \widehat{\mathrm{MV}}_{i,T_{i,n}}\right) + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i} T_{i,n}T_{j,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_{j,T_{j,n}}\right)^2\right] \\
&= \sum_{i=2}^{K}\mathbb{E}\left[\mathbb{E}\left[T_{i,n}\left(\mathrm{MV}_1 - \widehat{\mathrm{MV}}_{i,T_{i,n}}\right)\,\Big|\, T_{i,n}\right]\right] + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[\mathbb{E}\left[T_{i,n}T_{j,n}\left(\widehat{\mu}_{i,T_{i,n}} - \widehat{\mu}_{j,T_{j,n}}\right)^2\,\Big|\, T_{i,n}, T_{j,n}\right]\right] \\
&= \sum_{i=2}^{K}\mathbb{E}\left[T_{i,n}\left(\rho\mu_1 - \sigma_i^2\frac{T_{i,n}-1}{T_{i,n}} - \rho\mu_i + \sigma_1^2\right)\right] + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[T_{i,n}T_{j,n}\left(\Gamma_{i,j}^2 + \frac{\sigma_i^2}{T_{i,n}} + \frac{\sigma_j^2}{T_{j,n}}\right)\right] \\
&= \sum_{i=2}^{K}\mathbb{E}\left[T_{i,n}\right]\Delta_i + \sum_{i=2}^{K}\sigma_i^2 + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[T_{i,n}T_{j,n}\Gamma_{i,j}^2\right] + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[\sigma_i^2 T_{j,n} + \sigma_j^2 T_{i,n}\right] \\
&= \sum_{i=2}^{K}\mathbb{E}\left[T_{i,n}\right]\Delta_i + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[T_{i,n}T_{j,n}\right]\Gamma_{i,j}^2 + 3\sum_{i=1}^{K}\sigma_i^2 - \frac{2}{n}\sum_{i=1}^{K}\sigma_i^2\mathbb{E}\left[T_{i,n}\right] \\
&\leq \sum_{i=2}^{K}\mathbb{E}\left[T_{i,n}\right]\Delta_i + \frac{1}{n}\sum_{i=1}^{K}\sum_{j\neq i}\mathbb{E}\left[T_{i,n}T_{j,n}\right]\Gamma_{i,j}^2 + 3\sum_{i=1}^{K}\sigma_i^2.
\end{aligned}$$

This completes the proof of Lemma 1.

S-2. Figures and more numerical results

In this section, we show more numerical results to validate our theoretical results in the main paper.

In Figures S-1, S-2 and S-3, we report the expected regret of BMVTS with different ρ. Here the arm distributions are Bernoulli's with success probabilities $(0.1, 0.2, 0.23, 0.27, 0.32, 0.32, 0.34, 0.41, 0.43, 0.54, 0.55, 0.56, 0.67, 0.71, 0.79)$. The regret is averaged over 500 runs with a fixed time horizon $n = 30000$. These figures clearly show that BMVTS outperform LCB algorithm.

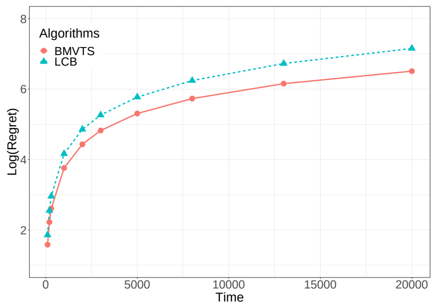

Figure S-1. Regrets for $\rho = 0.111$

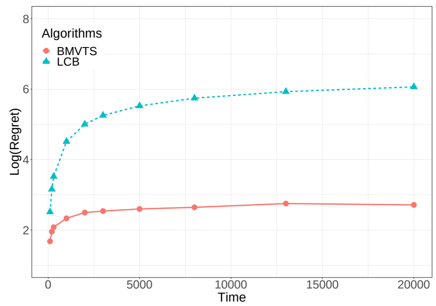

Figure S-2. Regrets for $\rho = 0.444$

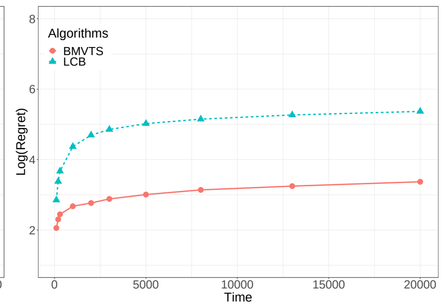

Figure S-3. Regrets for $\rho = 0.889$

S-3. Proof of Theorem 3

Since Theorem 3 is the most involved, we present it before Theorems 1 and 2. The proofs of the latter two theorems reuse several calculations that are done for the proof of Theorem 3.

S-3.1. Notations

We remind the reader of the definitions of the event $E_i(t)$ and the probability G_{is} as follows:

$$E_i\left(t\right) = \left\{\widehat{\mathrm{MV}}_{i,t} \leq \mathrm{MV}_1 - (1+\rho)\varepsilon\right\}, \quad G_{is} = \mathbb{P}_t\left(E_i(t)^c | T_{i,t} = s\right).$$

According to Lemma 2, we need to provide an upper bound for $\mathbb{E}[T_{i,n}]$.

S-3.2. Proofs of the lemmas

Before we get into the details of the proofs, let us present the proof of the lemmas in the main text and some other useful lemmas.

Lemma S-1 (Lemma 4 in the main text) *For a Gamma random variable $X \sim \mathrm{Gamma}\left(\alpha, \beta\right)$ with shape $\alpha \geq 2$ and rate $\beta > 0$, we have the following lower bound on the complementary cumulative distribution function*

$$\mathbb{P}\left(X \geq x\right) \geq \frac{1}{\Gamma(\alpha)} \exp\left(-\beta x\right)\left(1+\beta x\right)^{\alpha-1}, \text{ for } x > 0.$$

Proof: Let Y be an exponential random variable with rate β. Consider,

$$\begin{aligned}
\mathbb{P}\left(X \geq x\right) &= \int_x^\infty \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)}\,\mathrm{d}t \\
&= \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \int_0^\infty (z+x)^{\alpha-1} e^{-\beta z}\,\mathrm{d}z \\
&= \frac{\beta^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \int_0^\infty (z+x)^{\alpha-1} \beta e^{-\beta z}\,\mathrm{d}z \\
&= \frac{\beta^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \mathbb{E}\Big(\left[Y+x\right]^{\alpha-1}\Big) \\
&\geq \frac{\beta^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \Big(\mathbb{E}[Y+x]\Big)^{\alpha-1} \qquad \text{(S-1)} \\
&= \frac{\beta^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \left(\frac{1}{\beta}+x\right)^{\alpha-1} \\
&= \frac{e^{-\beta x}}{\Gamma(\alpha)} \left(1+\beta x\right)^{\alpha-1}
\end{aligned}$$

where (S-1) follows from Jensen's inequality and the convexity of $z \mapsto z^{\alpha-1}$ (recall that $\alpha \geq 2$).

Lemma S-2 (Harremoës (2016)) *Under the same setting as Lemma S-1,*

$$\mathbb{P}\left(X \geq x\right) \leq \exp\left(-2\alpha h\left(\frac{\beta x}{\alpha}\right)\right), \quad \text{for } x > \frac{\alpha}{\beta}.$$

where $h(x) = (x - 1 - \log x)/2$.

In the proofs of the following two lemmas, we use the following fact: Given $T_{i,t} = s, \widehat{\mu}_{i,s} = \mu < \mu_1, \widehat{\sigma}^2_{i,s} = \sigma^2 > \sigma^2_1$, $\theta_{i,t}$ and $\tau_{i,t}$ are independent because we sample them from different distributions independently.

Lemma S-3 (Tail Upper Bound) *We have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, T_{i,t} = s, \widehat{\mu}_{i,s} = \mu, \widehat{\sigma}^2_{i,s} = \sigma^2\right) \leq \exp\left(-\frac{s}{2}\left(\mu_1 - \mu - \varepsilon\right)^2\right) + \exp\left(-sh\left(\frac{\sigma^2}{\sigma^2_1 + \varepsilon}\right)\right).$$

Proof: We can compute this probability directly,

$$\begin{aligned}
&\mathbb{P}_t(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,|\, T_{i,t} = s, \widehat{\mu}_{i,s} = \mu, \widehat{\sigma}^2_{i,s} = \sigma^2) \\
&= \mathbb{P}_t\left(\rho\theta_{i,t} - \frac{1}{\tau_{i,t}} \geq \rho\mu_1 - \sigma^2_1 - (1+\rho)\varepsilon\right) \\
&= \mathbb{P}_t\left(\rho(\theta_{i,t} - \mu_1) + \left(\sigma^2_1 - \frac{1}{\tau_{i,t}}\right) \geq -(1+\rho)\varepsilon\right) \\
&\leq \mathbb{P}_t\left(\theta_{i,t} - \mu_1 \geq -\varepsilon\right) + \mathbb{P}_t\left(\frac{1}{\tau_{i,t}} - \sigma^2_1 \leq \varepsilon\right) \\
&\leq \exp\left(-\frac{s}{2}\left(\mu_1 - \mu - \varepsilon\right)^2\right) + \mathbb{P}_t\left(\tau_{i,t} \geq \frac{1}{\sigma^2_1 + \varepsilon}\right) \\
&\leq \exp\left(-\frac{s}{2}\left(\mu_1 - \mu - \varepsilon\right)^2\right) + \exp\left(-sh\left(\frac{\sigma^2}{\sigma^2_1 + \varepsilon}\right)\right).
\end{aligned}$$

The last inequality follows from Lemma S-2. The bound in Lemma S-3 is crucial for proving an upper bound of G_{is}, which is presented in Section S-3.4.

Lemma S-4 (Lemma 3 in the main text) *We have*

$$\begin{aligned}
&\mathbb{P}_t\big(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\big|\, T_{1,t} = s, \widehat{\mu}_{1,s} = \mu, \widehat{\sigma}^2_{1,s} = \sigma^2\big) \\
&\geq \begin{cases}
\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma^2_1 \leq \varepsilon\right)\mathbb{P}_t\left(\theta_{1,t} - \mu_1 \geq -\varepsilon\right) & \\
 & \text{if } \sigma^2 \geq \sigma^2_1, \mu \leq \mu_1 \\
\frac{1}{2}\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma^2_1 \leq \varepsilon\right) & \text{if } \sigma^2 \geq \sigma^2_1, \mu > \mu_1 \\
\frac{1}{2}\mathbb{P}_t\left(\theta_{1,t} - \mu_1 \geq -\varepsilon\right) & \text{if } \sigma^2 < \sigma^2_1, \mu \leq \mu_1 \\
\frac{1}{4} & \text{if } \sigma^2 < \sigma^2_1, \mu > \mu_1
\end{cases}.
\end{aligned} \tag{S-2}$$

Proof: Consider the following set of equalities and inequality,

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, T_{1,t} = s, \widehat{\mu}_{1,s} = \mu, \widehat{\sigma}^2_{1,s} = \sigma^2\right) \\
&= \mathbb{P}_t\left(\rho\theta_{1,t} - \frac{1}{\tau_{1,t}} \geq \rho\mu_1 - \sigma^2_1 - (1+\rho)\varepsilon\right) \\
&= \mathbb{P}_t\left(\rho(\theta_{1,t} - \mu_1) - \left(\frac{1}{\tau_{1,t}} - \sigma^2_1\right) \geq -(1+\rho)\varepsilon\right) \\
&\geq \mathbb{P}_t\left(\theta_{1,t} - \mu_1 \geq -\varepsilon\right) \cdot \mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma^2_1 \leq \varepsilon\right).
\end{aligned} \tag{S-3}$$

Then the lemma is proved by the inequality in (S-3), and

$$\mathbb{P}_t\left(\theta_{1,t} - \mu_1 \geq -\varepsilon\right) > \frac{1}{2} \quad \text{if} \quad \mu > \mu_1 \tag{S-4}$$

and

$$\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon\right) \geq \frac{1}{2} \quad \text{if} \quad \sigma^2 < \sigma_1^2. \tag{S-5}$$

Note that (S-4) and (S-5) can be established by using the properties of the median of Gaussian and Gamma distributions respectively.
Lemma S-4 provides us with a lower bound on G_{1s}, which is useful when we prove an upper bound of $\mathbb{E}\left[\frac{1}{G_{1s}} - 1\right]$ in Section S-3.3.

S-3.3. Bounding the first term of (16)

We now provide a bound for the first term of (16) in Lemma 2.
Let $c_1 = 1/\sqrt{2\pi\sigma_1^2}, c_2 = \frac{1}{2^{s/2}\Gamma(s/2)\sigma_1^{s-2}}, \tau = s(\sigma_1^2 + \varepsilon)$ and fix $\varepsilon > 0$. We define the conditional version of G_{1s} as

$$\widetilde{G}_{1s} = G_{1s}|_{\widehat{\mu}_{1,s}=\mu, \widehat{\sigma}_{1,s}^2=\beta} = \mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \widehat{\mu}_{1,s} = \mu, \widehat{\sigma}_{1,s}^2 = \beta\right)$$

which is the left-hand-side of (S-2) in Lemma S-4. Then we calculate the expectation $\mathbb{E}\left[\frac{1}{G_{1s}} - 1\right]$ by conditioning on various values of $\widehat{\mu}_{i,s}$ and $\widehat{\sigma}_{i,s}^2$. Note that we assumed that $\sigma_i^2 \leq 1$ for all $i = 1, \ldots, K$. For clarity, we partition the parameter space $(\beta, \mu) \in [0, \infty) \times (-\infty, \infty)$ into four parts as follows

$$[0, \infty] \times (-\infty, \infty) = A \cup B \cup C \cup D$$

where

$$\begin{aligned} A &= [0, \tau) \times [\mu_1 - \varepsilon, \infty), \quad & B &= [0, \tau) \times (-\infty, \mu_1 - \varepsilon], \\ C &= [\tau, \infty) \times [\mu_1 - \varepsilon, \infty), \quad & D &= [\tau, \infty) \times (-\infty, \mu_1 - \varepsilon]. \end{aligned}$$

Then the expectation of $(1/G_{1s}) - 1$ can be partitioned into four parts as follows,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{G_{1s}} - 1\right] &= c_1 c_2 \int_0^\infty \int_{-\infty}^\infty \frac{1 - \widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu - \mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\ &= c_1 c_2 \left(\int_A + \int_B + \int_C + \int_D\right) \frac{1 - \widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu - \mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \end{aligned} \tag{S-6}$$

Part A: Using the fourth case in Lemma S-4, we have

$$\frac{1 - \widetilde{G}_{1s}}{\widetilde{G}_{1s}} \geq 4(1 - \widetilde{G}_{1s}).$$

Then

$$
\begin{aligned}
& c_1 c_2 \int_A \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \le 4 c_1 c_2 \int_A (1-\widetilde{G}_{1s}) \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \le 4 c_1 c_2 \int_A \left(\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \le -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right) + \mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \ge \varepsilon \,\middle|\, \widehat{\sigma}_{1,s}^2 = \beta\right)\right) \\
& \qquad \cdot \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \le 4 c_1 \int_{\mu_1}^{\infty} \mathbb{P}_t\left(\theta_{1,t}-\mu_1 \le -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right) \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \\
& \qquad + 4 c_2 \int_{\mu_1}^{\infty} \mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \ge \varepsilon \,\middle|\, \widehat{\sigma}_{1,s}^2 = \beta\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\beta \\
& \le \; 4 c_1 \int_{\mu_1}^{\infty} \exp\left(-\frac{s(\mu-\mu_1+\varepsilon)^2}{2}\right) \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \\
& \qquad + 4 c_2 \int_0^{\tau} \exp\left(-\frac{\left(\beta - s\left(\sigma_1^2+\varepsilon\right)\right)^2}{4s\left(\sigma_1^2+\varepsilon\right)^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\beta && \text{(S-7)} \\
& \le \frac{8c_1}{s\varepsilon} \exp\left(-\frac{s\varepsilon^2}{2}\right) + 4 \exp\left(-\frac{s\varepsilon^2}{4(\sigma_1^2+\varepsilon)^2}\right) && \text{(S-8)} \\
& \le C_1 \exp\left(-\frac{s\varepsilon^2}{4}\right)
\end{aligned}
$$

where (S-7) follows from using tail upper bounds on the Gaussian and Gamma distributions. Here, and in the following, we use the notation $C_i, i \in \mathbb{N}$ to denote constants.

Part B: Using the third case in Lemma S-4, we have

$$
\frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \ge \frac{2}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \ge -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right)}.
$$

Then

$$
\begin{aligned}
& c_1 c_2 \int_B \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \le 2 c_1 c_2 \int_B \frac{1}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \ge -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \le 2 c_1 \int_{-\infty}^{\mu_1-\varepsilon} \frac{1}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \ge -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \cdot c_2 \int_0^{\tau} \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \mathrm{d}\beta \\
& \le 2 c_1 \int_{-\infty}^{\mu_1-\varepsilon} \frac{1}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \ge -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \\
& \le 2 c_1 \int_{-\infty}^{\mu_1-\varepsilon} \left(\sqrt{s}(\mu_1-\mu-\varepsilon) + \sqrt{s(\mu_1-\mu-\varepsilon)^2+4}\right) \exp\left(\frac{s(\mu-\mu_1+\varepsilon)^2}{2}\right) \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu && \text{(S-9)} \\
& \le 2 c_1 \int_0^{\infty} \left(\sqrt{s}z + \sqrt{sz^2+4}\right) \exp\left(\frac{sz^2}{2}\right) \exp\left(-\frac{s(z+\varepsilon)^2}{2}\right) \mathrm{d}z \\
& \le 2 c_1 \exp\left(-\frac{s\varepsilon^2}{2}\right) \int_0^{\infty} C_2 \sqrt{s} z \exp\left(-sz\varepsilon\right) \mathrm{d}z \\
& \le C_3 \exp\left(-\frac{s\varepsilon^2}{2}\right). && \text{(S-10)}
\end{aligned}
$$

For (S-9), we use the following well-known lower bound for the tail of Gaussian distribution (see for example, Formula 7.1.13 in Abramowitz & Stegun (1965)). Namely, for a Gaussian random variable X with mean μ and variance σ^2, we have

$$\mathbb{P}(X \geq \mu + \sigma x) \geq \sqrt{\frac{2}{\pi}} \cdot \frac{1}{x + \sqrt{x^2+4}} \exp\left(-\frac{x^2}{2}\right), \quad \forall x \geq 0.$$

For (S-10), we used integration by parts.

Part C: Use the second case in Lemma S-4, we have

$$\frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \geq \frac{2}{\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s} = \beta\right)}.$$

Then, we have

$$\begin{aligned}
& c_1 c_2 \int_C \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
\leq & \; 2c_1 c_2 \int_C \frac{1}{\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s} = \beta\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
\leq & \; c_1 \int_{\mu_1}^{\infty} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \cdot 2c_2 \int_{\tau}^{\infty} \frac{1}{\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s} = \beta\right)} \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\beta \\
\leq & \; 2c_2 \int_{\tau}^{\infty} \frac{1}{\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s} = \beta\right)} \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\beta \\
\leq & \; 2c_2 \Gamma\left(\frac{s}{2}\right) \int_{\tau}^{\infty} \exp\left(\frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)} - \frac{\beta}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} \left(1 + \frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)}\right)^{-\left(\frac{s}{2}-1\right)} \mathrm{d}\beta \qquad \text{(S-11)} \\
\leq & \int_{\tau}^{\infty} \exp\left(-\frac{\beta\varepsilon}{2\left(\sigma_1^2+\varepsilon\right)\sigma_1^2}\right) \left(\frac{\beta}{2\sigma_1^2}\right)^{\frac{s}{2}-1} \left(1 + \frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)}\right)^{-\left(\frac{s}{2}-1\right)} \mathrm{d}\beta \\
\leq & \; \left(\sigma_1^2+\varepsilon\right) \int_{s}^{\infty} \exp\left(-\frac{y\varepsilon}{2\sigma_1^2}\right) \left(\frac{\left(\sigma_1^2+\varepsilon\right) y}{2\sigma_1^2+\sigma_1^2 y}\right)^{\frac{s}{2}-1} \mathrm{d}y \\
\leq & \; C_4 \exp\left(-\frac{s\varepsilon}{2}\right)
\end{aligned}$$

where (S-11) follows from Lemma S-1.

Part D: Use the first case in Lemma S-4, we have

$$\frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \geq \frac{4}{\mathbb{P}_t\left(\theta_{1,t} - \mu_1 \geq -\varepsilon | \widehat{\mu}_{1,s} = \mu\right) \mathbb{P}_t\left(\frac{1}{\tau_{1,t}} - \sigma_1^2 \leq \varepsilon | \widehat{\sigma}_{i,s} = \beta\right)}.$$

Then

$$
\begin{aligned}
& c_1 c_2 \int_D \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \leq 4 c_1 c_2 \int_D \frac{1}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \geq -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right) \mathbb{P}_t\left(\frac{1}{\tau_{1,t}}-\sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s}=\beta\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\mu \,\mathrm{d}\beta \\
& \leq 2 c_1 \int_{-\infty}^{\mu_1} \frac{1}{\mathbb{P}_t\left(\theta_{1,t}-\mu_1 \geq -\varepsilon \,\middle|\, \widehat{\mu}_{1,s}=\mu\right)} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \\
& \quad \cdot 2 c_2 \int_{\tau}^{\infty} \frac{1}{\mathbb{P}_t\left(\frac{1}{\tau_{1,t}}-\sigma_1^2 \leq \varepsilon \,\middle|\, \widehat{\sigma}_{i,s}=\beta\right)} \beta^{\frac{s}{2}-1} e^{-\frac{\beta}{2\sigma_1^2}} \,\mathrm{d}\beta \\
& \leq C_5 \exp\left(-\frac{s\varepsilon^2}{2}-\frac{s\varepsilon}{2}\right).
\end{aligned} \tag{S-12}
$$

For (S-12), we can reuse the integrations in *Part B* and *Part C*.

Combine these four parts, we obtain an upper bound of (S-6) as follows,

$$
\mathbb{E}\left[\frac{1}{G_{1s}}-1\right] \leq C_1 \exp\left(-\frac{s\varepsilon^2}{4}\right)+C_3 \exp\left(-\frac{s\varepsilon^2}{2}\right)+C_4 \exp\left(-\frac{s\varepsilon}{2}\right)+C_5 \exp\left(-\frac{s\varepsilon^2}{2}-\frac{s\varepsilon}{2}\right)
$$

Summing over s, we have

$$
\sum_{s=1}^{\infty} \mathbb{E}\left[\frac{1}{G_{1s}}-1\right] \leq \frac{C_6}{\varepsilon^2}+\frac{C_7}{\varepsilon}+C_8
$$

S-3.4. Bounding the second term of (16)

Following from Lemma S-3, we have the following inclusions:

$$
\left\{\widehat{\mu}_{is}+\sqrt{\frac{2\log(2n)}{s}} \leq \mu_1-\varepsilon\right\} \subseteq \left\{\exp\left(-s\left(\Gamma_i-\varepsilon\right)^2\right) \leq \frac{1}{2n}\right\}
$$

and

$$
\left\{\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon} \geq h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right\} \cup \left\{\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon} \leq h_-^{-1}\left(\frac{\log(2n)}{s}\right)\right\} \subseteq \left\{\exp\left(-sh\left(\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon}\right)\right) \leq \frac{1}{2n}\right\}
$$

where $h_+^{-1}(y)=\max\{x: h(x)=y\}$, and $h_-^{-1}(y)=\min\{x: h(x)=y\}$.

Hence for

$$
s \geq u=\max\left\{\frac{2\log(2n)}{\left(\Gamma_i-2\varepsilon\right)^2}, \frac{\log(2n)}{h(\sigma_i^2/\sigma_1^2)}\right\},
$$

we have

$$
\begin{aligned}
\mathbb{P}_t\left(G_{is}>\frac{1}{n}\right) & \leq \mathbb{P}_t\left(\widehat{\mu}_{is}+\sqrt{\frac{2\log(2n)}{s}} \geq \mu_1-\varepsilon\right)+\mathbb{P}_t\left(h_-^{-1}\left(\frac{\log(2n)}{s}\right) \leq \frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon} \leq h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right) \\
& \leq \mathbb{P}_t\left(\widehat{\mu}_{i,s}-\mu_i \geq \Gamma_i-\varepsilon-\sqrt{\frac{2\log(2n)}{s}}\right)+\mathbb{P}_t\left(\widehat{\sigma}_i^2 \leq \left(\sigma_1^2+\varepsilon\right) h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right) \\
& \leq \exp\left(-\frac{s\left(\Gamma_i-\varepsilon-\sqrt{\frac{2\log(2n)}{s}}\right)^2}{2\sigma_i^2}\right)+\exp\left(-(s-1)\frac{\left(\left(\sigma_1^2+\varepsilon\right) h_+^{-1}\left(\frac{\log(2n)}{s}\right)-\sigma_i^2\right)^2}{4\sigma_i^4}\right) \\
& \leq \exp\left(-\frac{s\varepsilon^2}{\sigma_i^2}\right)+\exp\left(-(s-1)\frac{\varepsilon^2}{\sigma_1^4}\right).
\end{aligned}
$$

Summing over s,

$$\sum_{s=1}^{n} \mathbb{P}_t\left(G_{is} \geq 1/n\right) \leq u + \sum_{s=\lceil u \rceil}^{n} \exp\left(-\frac{s\varepsilon^2}{\sigma_i^2}\right) + \exp\left(-(s-1)\frac{\varepsilon^2}{\sigma_1^4}\right)$$
$$\leq 1 + \max\left\{\frac{2\log(2n)}{\left(\Gamma_i - 2\varepsilon\right)^2}, \frac{\log(2n)}{h(\frac{\sigma_i^2}{\sigma_1^2})}\right\} + \frac{2}{\varepsilon^2}.$$

Combining the two previous bounds, we have the following lemma,

Lemma S-5 *We have*

$$\mathbb{E}[T_{i,n}] \leq 1 + \max\left\{\frac{2\log(2n)}{\left(\Gamma_i - 2\varepsilon\right)^2}, \frac{\log(2n)}{h(\frac{\sigma_i^2}{\sigma_1^2})}\right\} + \frac{C_6}{\varepsilon^2} + \frac{C_7}{\varepsilon} + C_8$$

The finite-time regret bound for MVTS follows from Lemma S-5 and equation (10) in main text.

Theorem S-1 *The finite-time expected regret of MVTS for mean-variance Gaussian bandits satisfies*

$$\mathbb{E}\big[\widetilde{\mathcal{R}}_n\left(\mathrm{MVTS}\right)\big] \leq \sum_{i=2}^{K}\Bigg(1 + \max\left\{\frac{2\log(2n)}{\left(\Gamma_i - 2(\log n)^{-1/4}\right)^2}, \frac{\log(2n)}{h(\frac{\sigma_i^2}{\sigma_1^2})}\right\}$$
$$+ C_6\left(\log n\right)^{1/2} + C_7\left(\log n\right)^{1/4} + C_8\Bigg)\left(\Delta_i + 2\Gamma_{i,\max}^2\right).$$

Let $\varepsilon = (\log n)^{-\frac{1}{4}}$ and $n \to +\infty$, the regret bound in Theorem 3 follows from Theorem S-1.

S-4. Proof of Theorem 1

The proof is similar to that for proof of Theorem 3. For Theorem 1 (MTS), we define following event and conditional probability,

$$E_i\left(t\right) = \left\{\widehat{\mathrm{MV}}_{i,t} = \rho\theta_{i,t} - \widehat{\sigma}_{i,T_{i,t}}^2 \leq \mathrm{MV}_1 - (1+\rho)\varepsilon\right\}, \quad G_{is} = \mathbb{P}_t\left(E_i(t)^c | T_{i,t} = s\right)$$

Lemma S-6 *we have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\Big|\, \hat{\mu}_i = \mu, T_{i,t} = s\right) \leq \exp\left(-\frac{s\left(\mathrm{MV}_1 - \rho\mu - (1+\rho)\varepsilon\right)^2}{2\rho^2}\right)$$

Proof: Consider,

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\Big|\, \hat{\mu}_i = \mu, T_{i,t} = s\right)\\
&= \mathbb{P}_t\left(\rho\theta_{i,t} - \widehat{\sigma}_{i,s}^2 \geq \rho\mu_1 - \sigma_1^2 - (1+\rho)\varepsilon \,\big|\, \hat{\mu}_i = \mu, T_{i,t} = s\right)\\
&\leq \mathbb{P}_t\left(\rho\theta_{i,t} \geq \rho\mu_1 - \sigma_1^2 - (1+\rho)\varepsilon \,\big|\, \hat{\mu}_i = \mu, T_{i,t} = s\right)\\
&\leq \exp\left(-\frac{s\left(\mathrm{MV}_1 - \rho\mu - (1+\rho)\varepsilon\right)^2}{2\rho^2}\right)
\end{aligned}$$

This lemma is used to bound the second term of (16) in Lemma 2. We also need a lower bound of G_{1s} to bound the first term of (16) in Lemma 2.

Lemma S-7 *We have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\mu}_1 = \mu, T_{1,t} = s\right) \geq \begin{cases} \frac{1}{2}\mathbb{P}_t(\theta_{1,t} \geq \mu_1 - \varepsilon | \hat{\mu}_i = \mu, T_{1,t} = s) & \text{if } \widehat{\mu}_{i,s} \leq \mu_1 \\ \frac{1}{4} & \text{if } \widehat{\mu}_{i,s} > \mu_1 \end{cases}.$$

Proof: By direct calculation,

$$\begin{aligned} &\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\mu}_1 = \mu, T_{1,t} = s\right) \\ &= \mathbb{P}_t\left(\rho\theta_{1,t} - \widehat{\sigma}^2_{1,s} \geq \rho\mu_1 - \sigma_1^2 - (1+\rho)\varepsilon \,\middle|\, \hat{\mu}_i = \mu, T_{1,t} = s\right) \\ &\geq \mathbb{P}_t\left(\theta_{1,t} \geq \mu_1 - \varepsilon \,\middle|\, \hat{\mu}_1 = \mu, T_{1,t} = s\right)\mathbb{P}_t\left(\widehat{\sigma}^2_{1,s} \leq \sigma_1^2 + \varepsilon\right) \\ &\geq \frac{1}{2}\mathbb{P}_t\left(\theta_{1,t} \geq \mu_1 - \varepsilon \,\middle|\, \hat{\mu}_i = \mu, T_{1,t} = s\right) \end{aligned} \tag{S-13}$$

Then Lemma S-7 is proved by the inequality in (S-13) and the following fact: For X being an Gaussian random variable with mean μ and variance σ^2, if $x' < \mu$

$$\Pr(X > x') \geq \frac{1}{2}$$

S-4.1. Bounding the first term of (16)

With Lemma S-6, Lemma S-7, we can now prove Theorem 1.

Let $c = 1/\sqrt{2\pi\sigma_1^2}$ and fix $\varepsilon > 0$. We will condition on $\hat{\mu}_{i,s}$ and use the same proof technique as that for Theorem 3. The parameter space $(-\infty, \infty)$ will be divided into two parts, $(-\infty,\infty) = A \cup B$ where

$$A = (-\infty, \mu_1 - \varepsilon), \quad \text{and} \quad B = [\mu_1 - \varepsilon, \infty).$$

We define the conditional version of G_{1s} as

$$\widetilde{G}_{1s} = G_{1s}|_{\widehat{\mu}_{1,s}=\mu, \widehat{\sigma}^2_{1,s}=\beta} = \mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \widehat{\mu}_{1,s} = \mu\right)$$

Consider,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{G_{1s}} - 1\right] &= c\int_{-\infty}^{\infty} \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}} \exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right) \mathrm{d}\mu \\ &\leq 4c\int_A \left(1-\widetilde{G}_{1s}\right)\exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right)\mathrm{d}\mu + 2c\int_B \frac{\exp\left(-\frac{s(\mu-\mu_1)^2}{2\sigma_1^2}\right)}{\mathbb{P}_t\left(\theta_{1,t} \geq \mu_1 - \varepsilon \,\middle|\, \hat{\mu}_i = \mu, T_{1,t} = s\right)}\mathrm{d}\mu \\ &\leq C_9 \exp\left(-s\varepsilon^4/2\right) \end{aligned}$$

We have computed the same integration in (S-10) and (S-9). Summing over s, we have

$$\sum_{s=1}^{\infty} \mathbb{E}\left[\frac{1}{G_{1s}} - 1\right] \leq \frac{4C_9}{\varepsilon^2}.$$

S-4.2. Bounding the second term of (16)

Following Lemma S-6,

$$\left\{\hat{\mu}_{is} + \sqrt{\frac{2\log n}{s}} \leq \frac{\mathrm{MV}_1 - (1+\rho)\varepsilon}{\rho}\right\} \subseteq \left\{G_{is} \leq \frac{1}{n}\right\}$$

Hence for $s \geq u = \frac{2\rho^2 \log n}{\left(\rho\Gamma_i - \sigma_1^2 - (1+\rho)\varepsilon\right)^2}$, we have

$$\begin{aligned}\mathbb{P}_t\left(G_{is} > \frac{1}{n}\right) &\leq \mathbb{P}_t\left(\hat{\mu}_{is} + \sqrt{\frac{2\log n}{s}} \geq \frac{\mathrm{MV}_1 - (1+\rho)\varepsilon}{\rho}\right)\\ &= \mathbb{P}_t\left(\hat{\mu}_{is} - \mu_i \geq \Gamma_i - \frac{\sigma_1^2 + (1+\rho)\varepsilon}{\rho} - \sqrt{\frac{2\log n}{s}}\right)\\ &\leq \exp\left(-\frac{s\left(\Gamma_i - \frac{\sigma_1^2+(1+\rho)\varepsilon}{\rho} - \sqrt{\frac{2\log n}{s}}\right)^2}{2\sigma_i^2}\right)\end{aligned}$$

Summing over s,

$$\begin{aligned}\sum_{s=1}^{n} \mathbb{P}_t\left(G_{is} \geq 1/n\right) &\leq u + \sum_{s=\lceil u \rceil}^{n} \exp\left(-\frac{s\left(\Gamma_i - \frac{\sigma_1^2+(1+\rho)\varepsilon}{\rho} - \sqrt{\frac{2\log n}{s}}\right)^2}{2\sigma_i^2}\right)\\ &\leq 1 + \frac{2\rho^2 \log n}{\left(\rho\Gamma_i - \sigma_1^2 - (1+\rho)\varepsilon\right)^2} + \frac{2\sigma_i^2}{\left(\Gamma_i - \frac{\sigma_1^2+(1+\rho)\varepsilon}{\rho}\right)^2}\left(\sqrt{\pi\sigma_1^2\log n} + 1\right)\end{aligned}$$

Combining the two previous bounds, we have the following lemma.

Lemma S-8 *If $\rho > \max\left\{\frac{\sigma_1^2}{\Gamma_i}, i = 1, 2, \cdots, K\right\}$, we have*

$$\mathbb{E}[T_i(n)] \leq \frac{2\rho^2 \log n}{\left(\rho\Gamma_i - \sigma_1^2 - (1+\rho)\varepsilon\right)^2} + \frac{2}{\sigma_i^2\left(\Gamma_i - \frac{\sigma_1^2+(1+\rho)\varepsilon}{\rho}\right)^2}\left(\sqrt{\pi\sigma_1^2\log n} + 1\right) + \frac{4C_9}{\varepsilon^2} + 2$$

The finite-time regret bound follows from Lemma S-8 and equation (10) in main text.

Theorem S-2 *The finite-time expected regret of MVTS for mean-variance Gaussian bandits satisfies*

$$\mathbb{E}\big[\widetilde{\mathcal{R}}_n(\mathrm{MVTS})\big] \leq \sum_{i=2}^{K}\left(\frac{2\rho^2\log n}{\left(\rho\Gamma_i - \sigma_1^2 - (1+\rho)(\log n)^{-1/4}\right)^2} + 4C_9(\log n)^{1/2} + 2\right)\left(\Delta_i + 2\Gamma_{i,\max}^2\right).$$

Let $\varepsilon = (\log n)^{-\frac{1}{4}}$ and $n \to +\infty$, the regret bound in Theorem 1 follows from Theorem S-2.

S-5. Proof of Theorem 2

This is also similar to the proof of Theorem 3. For Theorem 2 (VTS), we define following event and conditional probability,

$$E_i(t) = \left\{\widehat{\mathrm{MV}}_{i,t} = \rho\widehat{\mu}_{i,s} - \frac{1}{\tau_{i,t}} \leq \mathrm{MV}_1 - (1+\rho)\varepsilon\right\}, \quad G_{is} = \mathbb{P}_t\left(E_i(t)^c | T_{i,t} = s\right).$$

Lemma S-9 *Given $\hat{\sigma}_{i,s}^2 = \sigma^2$ and $T_{i,t} = s$ such that*

$$s > \frac{2\sigma_i^2\log(2n)}{(\Gamma_{1,i} - \varepsilon)^2},$$

we have

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma_1^2, T_{i,t} = s\right) \leq \frac{1}{2n} + \exp\left(-sh\left(\frac{\sigma^2}{\sigma_1^2 + \varepsilon}\right)\right)$$

Proof: Consider,

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{i,t} = s\right) \\
&= \mathbb{P}_t\left(\rho\widehat{\mu}_{i,s} - \frac{1}{\tau_{i,t}} \geq \rho\mu_1 - \sigma_1^2 - (1+\rho)\varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{i,t} = s\right) \\
&\leq \mathbb{P}_t\left(\rho\widehat{\mu}_{i,s} \geq \rho\mu_1 - \rho\varepsilon\right) + \mathbb{P}_t\left(\frac{1}{\tau_{i,t}} \leq \sigma_1^2 + \varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{i,t} = s\right) \\
&\leq \frac{1}{2n} + \exp\left(-sh\left(\frac{\sigma^2}{\sigma_1^2 + \varepsilon}\right)\right).
\end{aligned}$$

This lemma is used to bound the second term of (16) in Lemma 2. We need also a lower bound of G_{1s} to bound the first term of (16) in Lemma 2.

Lemma S-10 *Given $\hat{\mu}_{1,s} = \mu$ and $T_{1,t} = s$, we have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{i,t} = s\right) \geq \begin{cases} \frac{1}{2}\mathbb{P}_t(\frac{1}{\tau_{1,t}} \leq \sigma_1^2 + \varepsilon \,|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{1,t} = s) & \text{if } \hat{\sigma}_{i,s}^2 \leq \sigma_1^2 \\ \frac{1}{4} & \text{if } \hat{\sigma}_{i,s}^2 > \sigma_1^2 \end{cases}.$$

Proof: By direct calculation,

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{1,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \hat{\mu}_1 = \mu, T_{1,t} = s\right) \\
&= \mathbb{P}_t\left(\rho\widehat{\mu}_{1,s} - \frac{1}{\tau_{1,t}} \geq \rho\mu_1 - \sigma_1^2 - (1+\rho)\varepsilon \,\middle|\, \hat{\sigma}_{1,s}^2 = \sigma^2, T_{1,t} = s\right) \\
&\geq \mathbb{P}_t\left(\frac{1}{\tau_{1,t}} \leq \sigma_1^2 + \varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{1,t} = s\right)\mathbb{P}_t\left(\widehat{\mu}_{1,s} \geq \mu_1 - \varepsilon\right) \\
&\geq \frac{1}{2}\mathbb{P}_t\left(\frac{1}{\tau_{1,t}} \leq \sigma_1^2 + \varepsilon \,\middle|\, \hat{\sigma}_{i,s}^2 = \sigma^2, T_{1,t} = s\right).
\end{aligned} \tag{S-14}$$

Then Lemma S-10 is proved by the inequality in (S-14) and the following fact: If X is an inverse-Gamma random variable with shape α and rate β, if $x > \frac{\beta}{\alpha-1}$

$$\Pr(X < x) \geq \frac{1}{2}.$$

S-5.1. Bounding the first term of (16)

Let $c = \frac{1}{2^{s/2}\Gamma(s/2)\sigma_1^{s-2}}$ and $\tau = s(\sigma_1^2 + \varepsilon)$ for some fixed $\varepsilon > 0$. To calculate the expectation conditioned on $\widehat{\sigma}_{i,s}^2$, we will use the same proof technique as the proof of Theorem 3. In particular, the parameter space $(0, \infty)$ will be divided into two parts, i.e., $(0, \infty) = A \cup B$ where

$$A = (0, \tau), \quad \text{and} \quad B = [\tau, \infty).$$

We define the conditional version of G_{1s} as

$$\widetilde{G}_{1s} = G_{1s}|_{\widehat{\mu}_{1,s}=\mu, \widehat{\sigma}_{1,s}^2=\beta} = \mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,t} \geq \mathrm{MV}_1 - (1+\rho)\varepsilon \,\middle|\, \widehat{\sigma}_{1,s}^2 = \beta\right)$$

Then

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{G_{1s}}-1\right] &= c\int_0^\infty \frac{1-\widetilde{G}_{1s}}{\widetilde{G}_{1s}}\beta^{\frac{s}{2}-1}e^{-\frac{\beta}{2\sigma_1^2}}\mathrm{d}\beta \\
&\le 4c\int_A (1-\widetilde{G}_{1s})\beta^{\frac{s}{2}-1}e^{-\frac{\beta}{2\sigma_1^2}}\mathrm{d}\beta + 2c\int_B \frac{1}{\mathbb{P}_t(\frac{1}{\tau_{1,t}}\le \sigma_1^2+\varepsilon|\hat{\sigma}_{i,s}^2=\sigma^2, T_{1,t}=s)}\beta^{\frac{s}{2}-1}e^{-\frac{\beta}{2\sigma_1^2}}\mathrm{d}\beta \\
&\le 4c\int_A \exp\left(-\frac{\left(\beta-s\left(\sigma_1^2+\varepsilon\right)\right)^2}{4s\left(\sigma_1^2+\varepsilon\right)^2}\right)\beta^{\frac{s}{2}-1}e^{-\frac{\beta}{2\sigma_1^2}}\mathrm{d}\beta \\
&\quad + 2c\Gamma(\frac{s}{2})\int_B \exp\left(\frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)}-\frac{\beta}{2\sigma_1^2}\right)\beta^{\frac{s}{2}-1}\left(1+\frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)}\right)^{-\left(\frac{s}{2}-1\right)}\mathrm{d}\beta \\
&\le 4\exp\left(-\frac{s\varepsilon^2}{4\sigma_1^2}\right)+\int_B \exp\left(-\frac{\beta\varepsilon}{2\left(\sigma_1^2+\varepsilon\right)\sigma_1^2}\right)\left(\frac{\beta}{2\sigma_1^2}\right)^{\frac{s}{2}-1}\left(1+\frac{\beta}{2\left(\sigma_1^2+\varepsilon\right)}\right)^{-\left(\frac{s}{2}-1\right)}\mathrm{d}\beta \\
&\le 4\exp\left(-\frac{s\varepsilon^2}{4\sigma_1^2}\right)+\left(\sigma_1^2+\varepsilon\right)\int_s^\infty \exp\left(-\frac{y\varepsilon}{2\sigma_1^2}\right)\left(\frac{\left(\sigma_1^2+\varepsilon\right)y}{2\sigma_1^2+\sigma_1^2 y}\right)^{\frac{s}{2}-1}\mathrm{d}\beta \\
&\le C_{10}\exp\left(-\frac{s\varepsilon^2}{4}\right)+C_{11}\exp\left(-\frac{s\varepsilon}{2}\right).
\end{aligned}
$$

We omit the details because the same integrations have been computed in (S-11) and (S-12). Summing from $s=0$ to ∞ shows that

$$
\sum_{s=0}^{\infty}\mathbb{E}\left[\frac{1}{G_{1s}}-1\right]\le \frac{4C_{10}}{\varepsilon^2}+\frac{2C_{11}}{\varepsilon}.
$$

S-5.2. Bounding the second term of (16)

Hence, similar to the analysis of MVTS,

$$
\left\{\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon}\ge h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right\}\cup\left\{\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon}\le h_-^{-1}\left(\frac{\log(2n)}{s}\right)\right\}\subseteq\left\{\exp\left(-sh\left(\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon}\right)\right)\le\frac{1}{2n}\right\}.
$$

Then if $s>\frac{\log(2n)}{h(\sigma_i^2/\sigma_1^2)}=s^*$, $\rho\le\frac{\Delta_i}{\Gamma_i}$, and $\Gamma_i^2\ge 8\sigma_1^2 h\left(\frac{\sigma_i^2}{\sigma_1^2}\right)$

$$
\begin{aligned}
\mathbb{P}_t\left(G_{is}>\frac{1}{n}\right) &\le \mathbb{P}_t\left(h_-^{-1}\left(\frac{\log(2n)}{s}\right)\le\frac{\widehat{\sigma}_i^2}{\sigma_1^2+\varepsilon}\le h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right) \\
&\le \mathbb{P}_t\left(\widehat{\sigma}_i^2\le\left(\sigma_1^2+\varepsilon\right)h_+^{-1}\left(\frac{\log(2n)}{s}\right)\right) \\
&\le \exp\left(-(s-1)\frac{\left(\left(\sigma_1^2+\varepsilon\right)h_+^{-1}\left(\frac{\log(2n)}{s}\right)-\sigma_i^2\right)^2}{4\sigma_i^4}\right) \\
&\le \exp\left(-(s-1)\frac{\varepsilon^2}{\sigma_1^4}\right).
\end{aligned}
$$

Hence

$$
\sum_{s=1}^{n}\mathbb{P}_t\left(G_{is}>\frac{1}{n}\right)\le s^*+2\sum_{s=\lceil s^*\rceil}^{n}\exp\left(-\left(s-1\right)\frac{\varepsilon^2}{\sigma_1^4}\right)=\frac{\log(2n)}{h\left(\frac{\sigma_i^2}{\sigma_1^2}\right)}+1+\frac{1}{\varepsilon^2}.
$$

Lemma S-11 *The number of times that VTS pulls arm i is bounded as*

$$
\mathbb{E}[T_{i,n}]\le\frac{\log(2n)}{h\left(\frac{\sigma_i^2}{\sigma_1^2}\right)}+\frac{4C_{10}}{\varepsilon^2}+\frac{2C_{11}}{\varepsilon}+C_{12}
$$

The finite-time regret bound follows from Lemma S-11 and equation (10) in main text,

Theorem S-3 *The finite-time expected regret of MVTS for mean-variance Gaussian bandits satisfies*

$$\mathbb{E}\big[\widetilde{\mathcal{R}}_n\,(\mathrm{MVTS})\,\big] \quad \leq \sum_{i=2}^{K}\left(\frac{\log(2n)}{h\left(\frac{\sigma_i^2}{\sigma_1^2}\right)} + 4C_{10}(\log n)^{1/2} + 2C_{11}(\log n)^{1/4} + C_{12}\right)\left(\Delta_i + 2\Gamma_{i,\max}^2\right).$$

Let $\varepsilon = (\log n)^{-\frac{1}{4}}$ and $n \to +\infty$, the regret bound in Theorem 2 follows from Theorem S-3.

S-6. Proof of Theorem 4

We provide following useful lemmas before we process to prove the theorem.

Lemma S-12 (Chernoff-Hoeffding bound I) *Let $X_1, \ldots, X_n$ be independent $\{0,1\}$-valued random variables (i.e., Bernoulli random variables) with $\mathbb{E}[X_i] = p_i$. Let $X = \frac{1}{n}\sum_{i=1}^{n} X_i, \mu = \mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} p_i$. Then, for any $0 < \lambda < 1-\mu$,*

$$\mathbb{P}(X \geq \mu + \lambda) \leq \exp(-nd(\mu+\lambda, \mu))$$

and, for any $0 < \lambda < \mu$,

$$\mathbb{P}(X \leq \mu - \lambda) \leq \exp(-nd(\mu-\lambda, \mu))$$

where $d(a,b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$.

Lemma S-13 (Chernoff-Hoeffding bound II) *Let $X_1, \cdots, X_n$ be random variables with common range $[0,1]$ and such that $\mathbb{E}[X_t|X_1, \ldots, X_{t-1}] = \mu$. Let $S_n = X_1 + \ldots + X_n$. Then for all $a \geq 0$,*

$$\mathbb{P}\left(S_n \geq n\mu + a\right) \leq e^{-2a^2/n},$$

$$\mathbb{P}\left(S_n \leq n\mu - a\right) \leq e^{-2a^2/n}.$$

Lemma S-14 (Relationship between Beta distribution and Binomial distribution) *For all positive integers α, β,*

$$F_{\alpha,\beta}^{\mathrm{Beta}}(y) = 1 - F_{\alpha+\beta-1,y}^{\mathrm{B}}(\alpha - 1).$$

Lemma S-15 *If M is a Binomial random variable with s trials and probability of success p, then*

$$\mathbb{P}\left(d(\tilde{p}-\varepsilon, M/s) \leq \frac{\log(2n)}{s}\right) \leq \exp\left(-2s\Big(\tilde{p} - p - \varepsilon \pm \sqrt{\frac{\log(2n)}{s}}\Big)^2\right).$$

where the $\pm$ is taken so that the exponent is minimized.

Proof: Let us define

$$\mathrm{p}_s := \mathbb{P}\left(d(\tilde{p}-\sqrt{\varepsilon}, M/s) \leq \frac{\log(2n)}{s}\right).$$

Clearly by the law of large numbers $\mathrm{p}_s \to 0$ as $s \to \infty$. We can write

$$\mathrm{p}_s := \mathbb{P}\left(d\Big(\tilde{p}-\sqrt{\varepsilon}, \frac{1}{s}\sum_{i=1}^{s} X_i\Big) \leq \frac{\log(2n)}{s}\right),$$

where X_i are i.i.d. Bernoulli random variables with probability of success p.

Now by a slightly strengthened form of Sanov's theorem (Csiszár & Körner, 2011, Problem 2.12(c)), we have the large deviations bound

$$\mathrm{p}_s \leq \exp\left(-s\min_{q\in\mathcal{A}_s} d(q, p_1)\right)$$

where

$$\mathcal{A}_s := \left\{ q \in [0,1] : d(p_1 - \sqrt{\varepsilon}, q) \leq \frac{2\log(2n)}{s} \right\}.$$

By Pinsker's inequality $d(p,q) \geq 2(p-q)^2$ (assuming natural logs) so

$$\mathcal{A}_s \subset \mathcal{A}'_s := \left\{ q \in [0,1] : (p_1 - \sqrt{\varepsilon} - q)^2 \leq \frac{\log(2n)}{s} \right\}.$$

Hence, one has

$$\mathrm{p}_s \leq \exp\left(-s \min_{q \in \mathcal{A}'_s} d(q, p_1) \right) = \exp\left(-sd\Big(p_1 - \sqrt{\varepsilon} \pm \sqrt{\frac{\log(2n)}{s}}, p_1\Big) \right).$$

The lemma is proven by applying Pinsker's inequality again.

Lemma S-16 (Tail Upper Bound) *We have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m \right) \leq \exp\left(-sd\left(x - \sqrt{\varepsilon}, \frac{m}{s} \right) \right) + \exp\left(-sd\left(y + \sqrt{\varepsilon}, \frac{m}{s} \right) \right).$$

Proof: Let $x = \frac{(1-\rho)+|1-\rho-2p_1|}{2}, y = \frac{(1-\rho)-|1-\rho-2p_1|}{2}$, then

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m \right) \\
&= \mathbb{P}_t\left(\rho\theta_{i,t} - \theta_{i,t}(1-\theta_{i,t}) \geq \rho p_1 - p_1(1-p_1) - \epsilon \,\big|\, T_{i,t} = s, \alpha_{i,t} = m \right) \\
&= \mathbb{P}_t\left(\theta_{i,t}^2 - p_1^2 - (1-p)(\theta_{i,t} - p_1) \geq -\epsilon \,\big|\, T_{i,t} = s, \alpha_{i,t} = m \right) \\
&= \mathbb{P}_t\left(\theta_{i,t} \geq \frac{(1-\rho) + \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2} \right) \\
&\qquad\qquad + \mathbb{P}_t\left(\theta_{i,t} \leq \frac{(1-\rho) - \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2} \right) \\
&\leq \mathbb{P}_t\left(\theta_{i,t} \geq \frac{(1-\rho) + |1-\rho-2p_1|}{2} - \sqrt{\varepsilon} \right) + \mathbb{P}_t\left(\theta_{i,t} \leq \frac{(1-\rho) - |1-\rho-2p_1|}{2} + \sqrt{\varepsilon} \right) \\
&\leq \exp\left(-sd\left(x - \sqrt{\varepsilon}, \frac{m}{s} \right) \right) + \exp\left(-sd\left(y + \sqrt{\varepsilon}, \frac{m}{s} \right) \right).
\end{aligned}$$

The last inequality follows from Chernoff-Hoeffding bound (Lemma S-12).

Lemma S-17 (Tail Lower Bound) *We have*

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m \right) \geq \begin{cases} F^{\mathrm{B}}_{s+1,\tilde{y}}(m) & \text{if } 1-\rho-2p_1 < 0 \\ 1 - F^{\mathrm{B}}_{s+1,\tilde{y}}(m) & \text{if } 1-\rho-2p_1 > 0 \end{cases}.$$

where

$$\tilde{x} = \frac{(1-\rho) + \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2}, \qquad \tilde{y} = \frac{(1-\rho) - \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2}$$

and $F^{\mathrm{B}}_{n,p}(\cdot)$ is the cumulative distribution function of the Binomial distribution.

Proof: Consider,

$$\begin{aligned}
&\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m\right) \\
&= \mathbb{P}_t(\rho\theta_{i,t} - \theta_{i,t}(1-\theta_{i,t}) \geq \rho p_1 - p_1(1-p_1) - \epsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m) \\
&= \mathbb{P}_t(\theta_{i,t}^2 - p_1^2 - (1-p)(\theta_{i,t} - p_1) \geq -\epsilon \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m) \\
&= \mathbb{P}_t\left(\theta_{i,t} \geq \frac{(1-\rho) + \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2} \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m\right) \\
&\qquad + \mathbb{P}_t\left(\theta_{i,t} \leq \frac{(1-\rho) - \sqrt{(1-\rho)^2 + 4(p_1^2 - p_1(1-\rho) - \varepsilon)}}{2} \,\Big|\, T_{i,t} = s, \alpha_{i,t} = m\right) \\
&= F^{\mathrm{B}}_{s+1,\tilde{x}}(m) + 1 - F^{\mathrm{B}}_{s+1,\tilde{y}}(m).
\end{aligned}$$

Then we have following lower bound,

$$\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, S_{i,t} = m\right) \geq \begin{cases} F^{\mathrm{B}}_{s+1,\tilde{x}}(m) & \text{if } 1-\rho-2p_1 < 0 \\ 1 - F^{\mathrm{B}}_{s+1,\tilde{y}}(m) & \text{if } 1-\rho-2p_1 > 0 \end{cases}.$$

S-6.1. Bounding the first term of (16)

With Lemma S-16, Lemma S-17, we can now prove Theorem 4.

Fix $\varepsilon > 0$. We will condition on $S_{i,s}$ and use the same proof technique as the proof of Theorem 3.

Consider,

$$\mathbb{E}\left[\frac{1}{G_{1s}} - 1\right] \leq \sum_{m=0}^{s} \frac{1}{\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, S_{i,t} = m\right)} \binom{s}{m} p_1^m (1-p_1)^{s-m}.$$

Case 1: If $1-\rho-2p_1 > 0$,

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{G_{1s}} - 1\right] &\leq \sum_{m=0}^{s} \frac{1}{\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s} \geq \mathrm{MV}_1 - \varepsilon \,\Big|\, T_{i,t} = s, S_{i,t} = m\right)} \binom{s}{m} p_1^m (1-p_1)^{s-m} \\
&\leq \sum_{m=0}^{s} \frac{1}{1 - F^{\mathrm{B}}_{s+1,\tilde{y}}(m)} \binom{s}{m} p_1^m (1-p_1)^{s-m} \\
&\leq \sum_{m=0}^{\lfloor \tilde{y}s \rfloor} 2\binom{s}{m} p_1^m (1-p_1)^{s-m} + \sum_{m=\lfloor \tilde{y}s \rfloor + 1}^{s} \frac{\binom{s}{m} p_1^m (1-p_1)^{s-m}}{\binom{s+1}{m} \tilde{y}^m (1-\tilde{y})^{s+1-m}} \\
&\leq 2\exp\left(-\frac{2(\lfloor \tilde{y}s \rfloor - sp_1)^2}{s}\right) + \frac{1}{1-\tilde{y}} \sum_{m=\lfloor \tilde{y}s \rfloor + 1}^{s} \frac{p_1^m (1-p_1)^{s-m}}{\tilde{y}^m (1-\tilde{y})^{s-m}} \\
&\leq 2\exp(-2s(\tilde{y} - p_1)^2) + \frac{p_1}{1-\tilde{y}} \exp(-sd(\tilde{y}, p_1)).
\end{aligned}$$

The first part follows from Lemma S-13, the second part is by direct computation.

Case 2: If $1-\rho-2p_1<0$,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{G_{1s}}-1\right] &\le \sum_{m=0}^{s}\frac{1}{\mathbb{P}_t\left(\widehat{\mathrm{MV}}_{i,s}\ge \mathrm{MV}_1-\varepsilon \,\Big|\, T_{i,t}=s,S_{i,t}=m\right)}\binom{s}{m}p_1^m(1-p_1)^{s-m}\\ &\le \sum_{m=0}^{s}\frac{1}{F^{\mathrm{B}}_{s+1,\tilde{x}}(m)}\binom{s}{m}p_1^m(1-p_1)^{s-m}\\ &\le \sum_{m=0}^{\lfloor\tilde{x}s\rfloor}\frac{\binom{s}{m}p_1^m(1-p_1)^{s-m}}{\binom{s+1}{m}\tilde{x}^m(1-\tilde{x})^{s+1-m}}+\sum_{m=\lfloor\tilde{x}s\rfloor+1}^{s}2\binom{s}{m}p_1^m(1-p_1)^{s-m}\\ &\le \frac{1}{1-\tilde{x}}\sum_{m=0}^{\lfloor\tilde{x}s\rfloor}\frac{p_1^m(1-p_1)^{s-m}}{\tilde{x}^m(1-\tilde{x})^{s-m}}+2\exp\Big(-\frac{2(\lfloor\tilde{x}s\rfloor+1-sp_1)^2}{s}\Big)\\ &\le \frac{p_1}{1-\tilde{x}}\exp(-sd(\tilde{x},p_1))+2\exp\Big(-\frac{2(\lfloor\tilde{x}s\rfloor+1-sp_1)^2}{s}\Big).\end{aligned}$$

The first part follows from Lemma S-13, the second part is by direct computation.

Summing over s, we have
Case 1:

$$\sum_{s=1}^{\infty}\mathbb{E}\left[\frac{1}{G_{1s}}-1\right]\le\frac{C_{13}}{(\tilde{y}-p_1)^2}+\frac{C_{14}}{d(\tilde{y},p_1)}. \tag{S-15}$$

Case 2:

$$\sum_{s=1}^{\infty}\mathbb{E}\left[\frac{1}{G_{1s}}-1\right]\le\frac{C_{15}}{(\tilde{x}-p_1)^2}+\frac{C_{16}}{d(\tilde{x},p_1)}. \tag{S-16}$$

S-6.2. Bounding the second term of (16)

Follow from Lemma S-16, we have the following inclusions:

$$\left\{d\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\ge\frac{\log(2n)}{s}\right\}\subseteq\left\{\exp\left(-sd\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\right)\le\frac{1}{2n}\right\}$$

and

$$\left\{d\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\ge\frac{\log(2n)}{s}\right\}\subseteq\left\{\exp\left(-sd\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\right)\le\frac{1}{2n}\right\}$$

Hence for

$$s\ge u=\max\left\{\frac{\log(2n)}{2\left(\Gamma_i-\sqrt{\varepsilon}\right)^2},\frac{\log(2n)}{2\left(1-\rho-p_1-p_i-\sqrt{\varepsilon}\right)^2}\right\},$$

we have

$$\mathbb{P}_t\left(G_{is}>\frac{1}{n}\right)\le\mathbb{P}_t\left(d\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\le\frac{\log(2n)}{s}\right)+\mathbb{P}_t\left(d\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\le\frac{\log(2n)}{s}\right). \tag{S-17}$$

Case 1: If $1-\rho-2p_1>0$, then $x=1-\rho-p_1>p_1$, $y=p_1$, then, the first term of (S-17) can be bounded by applying Lemma S-15,

$$\begin{aligned}&\mathbb{P}_t\left(d\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\le\frac{\log(2n)}{s}\right)\\ =\;&\mathbb{P}_t\left(d\left(1-\rho-p_1-\sqrt{\varepsilon},\frac{m}{s}\right)\le\frac{\log(2n)}{s}\right)\\ \le\;&\exp\left(-2s\Big(1-\rho-p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right).\end{aligned}$$

The second term of (S-17) is bounded as follows,

$$\begin{aligned}
&\mathbb{P}_t\left(d\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&=\mathbb{P}_t\left(d\left(p_1-\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&\leq\ \exp\left(-2s(p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}})^2\right).
\end{aligned}$$

Case 2: If $1-\rho-2p_1<0$, then $x=p_1$, $y=1-\rho-p_1<p_1$, we then apply Lemma S-15 again,

$$\begin{aligned}
&\mathbb{P}_t\left(d\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&=\mathbb{P}_t\left(d\left(p_1-\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&\leq\ \exp\left(-2s\Big(p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right).
\end{aligned}$$

The second term of (S-17) is bounded as follows,

$$\begin{aligned}
&\mathbb{P}_t\left(d\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&=\mathbb{P}_t\left(d\left(1-\rho-p_1-\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&\leq\ \exp\left(-2s\Big(1-\rho-p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right).
\end{aligned}$$

Combining the two cases, we have

$$\begin{aligned}
\mathbb{P}_t\left(G_{is}>\frac{1}{n}\right)&\leq\mathbb{P}_t\left(d\left(x-\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)+\mathbb{P}_t\left(d\left(y+\sqrt{\varepsilon},\frac{m}{s}\right)\leq\frac{\log(2n)}{s}\right)\\
&\leq\ \exp\left(-2s\Big(p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right)+\exp\left(-2s\Big(1-\rho-p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right).
\end{aligned}$$

Summing over s,

$$\begin{aligned}
\sum_{s=1}^{n}\mathbb{P}_t\left(G_{is}\geq 1/n\right)&\leq u+\sum_{s=\lceil u\rceil}^{n}\exp\left(-2s\Big(p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right)\\
&\qquad+\exp\left(-2s\Big(1-\rho-p_1-p_i-\sqrt{\varepsilon}\pm\sqrt{\frac{\log(2n)}{2s}}\Big)^2\right)\\
&\leq\max\left\{\frac{\log(2n)}{2\left(\Gamma_i-\sqrt{\varepsilon}\right)^2},\frac{\log(2n)}{2\left(1-\rho-p_1-p_i-\sqrt{\varepsilon}\right)^2}\right\}+\frac{C_{17}}{\varepsilon^2}+\frac{C_{18}}{\varepsilon}. \qquad \text{(S-18)}
\end{aligned}$$

Lemma S-18 *The expected number of times that BMVTS pulls arm i is bounded as*

$$\mathbb{E}[T_{i,n}]\leq\max\left\{\frac{\log(2n)}{2\left(\Gamma_i-\sqrt{\varepsilon}\right)^2},\frac{\log(2n)}{2\left(1-\rho-p_1-p_i-\sqrt{\varepsilon}\right)^2}\right\}+\frac{C_{19}}{\varepsilon^2}+\frac{C_{20}}{\varepsilon}.$$

Proof: This lemma follows from equations (S-15), (S-16), (S-18), and notice that as $\varepsilon \to 0^+$,

$$\frac{1}{(\tilde{x}-p_1)^2} = \Theta\Big(\frac{1}{\varepsilon^2}\Big), \quad \frac{1}{(\tilde{y}-p_1)^2} = \Theta\Big(\frac{1}{\varepsilon^2}\Big), \quad \frac{1}{d(\tilde{x},p_1)} = \Theta\Big(\frac{1}{\varepsilon^2}\Big), \quad \frac{1}{d(\tilde{y},p_1)} = \Theta\Big(\frac{1}{\varepsilon^2}\Big)$$

The finite-time regret bound follows from Lemma S-18 and equation (10) in the main text,

Theorem S-4 *The finite-time expected regret of BMVTS for mean-variance Bernoulli bandits satisfies*

$$\mathbb{E}\big[\widetilde{\mathcal{R}}_n\,(\mathrm{BMVTS})\,\big] \le \sum_{i=2}^{K}\left(\max\left\{\frac{\log(2n)}{2\left(\Gamma_i-\sqrt{\varepsilon}\right)^2}, \frac{\log(2n)}{2\left(1-\rho-p_1-p_i-\sqrt{\varepsilon}\right)^2}\right\} + \frac{C_{19}}{\varepsilon^2} + \frac{C_{20}}{\varepsilon}\right)\left(\Delta_i + 2\Gamma_{i,\max}^2\right).$$

Let $\varepsilon = (\log n)^{-\frac{1}{4}}$ and $n \to +\infty$, the regret bound in Theorem 4 follows from Theorem S-4.