Online Appendix

Diffusion Approximations for a Class of Sequential Experimentation Problems

Victor Araman: Olayan School of Business, American University of Beirut, Beirut, Lebanon.
René Caldentey: Booth School of Business, The University of Chicago.

A Proofs

Proof of Lemma 1: Let t_i be the i^{th} jump of N_t and let δ_{t_i-} be the decision maker's belief just before observing the outcome x_{t_i} of experiment $\mathcal{E}_{t_i}$. Then, by Bayes's rule we have that

$$\delta_{t_i} = \mathbb{P}(\Theta = \theta_0 | \mathcal{F}_{t_i-}, x_{t_i}) = \frac{\mathbb{P}(x_{t_i-}|\mathcal{F}_{t_i-}, \Theta = \theta_0)\,\mathbb{P}(\Theta = \theta_0|\mathcal{F}_{t_i-})}{\mathbb{P}(x_{t_i}|\mathcal{F}_{t_i-})} = \frac{Q(x_{t_i}, \mathcal{E}_{t_i}, \theta_0)\,\delta_{t_i-}}{Q(x_{t_i}, \mathcal{E}_{t_i}, \theta_0)\,\delta_{t_i-} + Q(x_{t_i}, \mathcal{E}_{t_i}, \theta_1)\,(1-\delta_{t_i-})}$$
$$= \frac{\delta_{t_i-}}{\delta_{t_i-} + (1-\delta_{t_i-})\,\mathcal{L}(x_{t_i}, \mathcal{E}_{t_i})}.$$

By iterating this recursion, with $\delta_{t_0} = \delta$, we get that

$$\delta_{t_i} = \frac{\delta}{\delta + (1-\delta)\,L_{t_i}}, \qquad \text{where } L_{t_i} = \prod_{j=1}^{i} \mathcal{L}(x_{t_j}, \mathcal{E}_{t_j}).$$

Finally, the result follows from noticing that δ_t is a pure jump process and so $\delta_t = \delta_{t_{N_t}}$. □

Proof of Proposition 1: First, the convexity of $G(\delta)$ follows directly from its representation in (1) and the fact that the 'max' of convex functions is also a convex function. To prove the convexity of the value function $\Pi(\delta)$ for $\delta \in (0,1)$, let us first recall that the value function $\Pi(\delta)$ satisfies the HJB equation:

$$\Pi(\delta) = \max\left\{G(\delta)\,,\ \frac{\Lambda}{\Lambda + r}\max_{\mathcal{E}\in\mathscr{E}}\left\{\mathbb{E}_\delta\Big[\Pi\big(\delta + \eta(\delta, x, \mathcal{E})\big)\Big]\right\}\right\}. \tag{A-1}$$

Now consider a sequence of functions $\{\Pi_k(\delta)\}_{k\geq 0}$ defined recursively by

$$\Pi_{k+1}(\delta) = \max\left\{G(\delta)\,,\ \frac{\Lambda}{\Lambda + r}\max_{\mathcal{E}\in\mathscr{E}}\left\{\mathbb{E}_\delta\Big[\Pi_k\big(\delta + \eta(\delta, x, \mathcal{E})\big)\Big]\right\}\right\}, \qquad k = 0, 1, \ldots$$

with $\Pi_0(\delta) = G(\delta)$. It is easy to see that the functions $\{\Pi_k(\delta)\}_{k\geq 0}$ are continuous in δ and pointwise monotonically increasing in k, that is, $\Pi_{k+1}(\delta) \geq \Pi_k(\delta)$ for all $\delta \in (0,1)$. Furthermore, the sequence converges uniformly to a limit $\Pi(\delta) := \lim_{k\to\infty}\Pi_k(\delta)$ that satisfies the HJB equation in (A-1). To see

this, note that

$$\begin{aligned}0 \le \Pi_{k+1}(\delta) - \Pi_k(\delta) &\le \frac{\Lambda}{\Lambda + r}\,\left[\max_{\mathcal{E}\in\mathscr{E}}\Big\{\mathbb{E}_\delta\Big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E})\big)\Big]\Big\} - \max_{\mathcal{E}\in\mathscr{E}}\Big\{\mathbb{E}_\delta\Big[\Pi_{k-1}\big(\delta+\eta(\delta,x,\mathcal{E})\big)\Big]\Big\}\right]\\ &\le \frac{\Lambda}{\Lambda + r}\,\Big[\mathbb{E}_\delta\Big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E}_k^*(\delta))\big)\Big] - \mathbb{E}_\delta\Big[\Pi_{k-1}\big(\delta+\eta(\delta,x,\mathcal{E}_k^*(\delta))\big)\Big]\Big]\\ &\le \frac{\Lambda}{\Lambda + r}\,\mathbb{E}_\delta\Big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E}_k^*(\delta))\big) - \Pi_{k-1}\big(\delta+\eta(\delta,x,\mathcal{E}_k^*(\delta))\big)\Big],\end{aligned}$$

where

$$\mathcal{E}_k^*(\delta) := \operatorname*{argmax}_{\mathcal{E}\in\mathscr{E}}\Big\{\mathbb{E}_\delta\Big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E})\big)\Big]\Big\}.$$

Taking the 'sup' over δ, it follows that

$$\rho_k \le \frac{\Lambda}{\Lambda + r}\,\rho_{k-1} \le \left(\frac{\Lambda}{\Lambda + r}\right)^k \sup_{\delta\in(0,1)}\Big\{G(\delta)\Big\}, \qquad \text{where } \rho_k := \sup_{\delta\in(0,1)}\Big\{\Pi_{k+1}(\delta) - \Pi_k(\delta)\Big\}$$

and so $\rho_k \to 0$ as $k \to \infty$.

We now complete the proof by showing that the HJB operator preserves convexity. That is, if $\Pi_k(\delta)$ is convex in $(0,1)$ then $\Pi_{k+1}(\delta)$ is also convex. First, since $G(\delta)$ is convex, it follows trivially that $\Pi_1(\delta)$ is convex. Now, let us suppose that $\Pi_k(\delta)$ is convex in $(0,1)$. Then, since the 'max' operator preserves convexity, we just need to show that $\mathbb{E}_\delta\big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E})\big)\big]$ is convex. We can rewrite this expectation as follows:

$$\mathbb{E}_\delta\Big[\Pi_k\big(\delta+\eta(\delta,x,\mathcal{E})\big)\Big] = \sum_{x\in\mathcal{X}_\mathcal{E}} \Pi_k\left(\frac{\delta\,Q(x,\mathcal{E},\theta_0)}{\delta\,Q(x,\mathcal{E},\theta_0) + (1-\delta)\,Q(x,\mathcal{E},\theta_1)}\right)\big(\delta\,Q(x,\mathcal{E},\theta_0) + (1-\delta)\,Q(x,\mathcal{E},\theta_1)\big).$$

Since the sum of convex functions is convex, we will show that each summand on the right-hand side above is convex. To ease notation, let us define $Q_0 = Q(x,\mathcal{E},\theta_0)$, $Q_1 = Q(x,\mathcal{E},\theta_1)$, $y = \delta\,Q_0 + (1-\delta)\,Q_1$ and define the function

$$H(y) := y\,\Pi_k\left(\frac{ay+b}{y}\right), \qquad \text{where } a := \frac{Q_0}{Q_0 - Q_1} \quad \text{and} \quad b := \frac{Q_0\,Q_1}{Q_1 - Q_0}.$$

Since y is a linear transformation of δ we can focus on proving the convexity of $H(y)$ for $y \in (Q_0, Q_1)$ (assuming, without loss of generality, that $Q_0 < Q_1$). To this end, we will use the following characterization of a convex function:

> Let a continuous function $h(\delta)$ be such that for any δ in the interior of the domain of h the subdifferential $\partial h(\delta)$ is not empty. Then h is convex. (see Theorem 3.2.6 in Bazaraa et al. 1993)

Thus, we would like to show that for every $y \in (Q_0, Q_1)$, there exists a subdifferential ∂H_y such that $H(z) \ge H(y) + \partial H_y\,(z-y)$, for all $z \in (Q_0, Q_1)$. Since $\Pi_k(\delta)$ is convex then for every $\delta \in (0,1)$ there exists a subdifferential $\partial\Pi_\delta$ such that $\Pi_k(z) \ge \Pi_k(\delta) + \partial\Pi_\delta\,(z-\delta)$, for all $z \in (0,1)$. For $y \in (Q_0, Q_1)$ define

$$\hat{y} := \frac{a\,y + b}{y}$$

then by the convexity of $\Pi_k(\delta)$ for any $z \in (Q_0, Q_1)$ we have

$$\Pi_k\left(\frac{a\,z+b}{z}\right) \geq \Pi_k\left(\frac{a\,y+b}{y}\right) + \partial\Pi_{\hat{y}}\left(\frac{a\,z+b}{z} - \frac{a\,y+b}{y}\right).$$

Multiplying by $z\,y$ (which is nonnegative since $Q_0 > 0$) and rearranging terms we get that

$$y\,H(z) \geq z\,H(y) - b\,\partial\Pi_{\hat{y}}\,(z-y) \quad \Longleftrightarrow \quad H(z) \geq H(y) + \left(\frac{H(y) - b\,\partial\Pi_{\hat{y}}}{y}\right)(z-y)$$

and so

$$\partial H_y := \left(\frac{H(y) - b\,\partial\Pi_{\hat{y}}}{y}\right)$$

is a subdifferential for H at y. This completes the proof. □

Proof of Lemma 2: Recall that the belief process can be written in terms of the likelihood function L as follows:

$$\delta_t = \frac{\delta}{\delta + (1-\delta)\,L_t}, \qquad \text{where} \quad L_t := \prod_{i=0}^{N_t} \mathcal{L}(x_{t_i}, \mathcal{E}_{t_i}).$$

Now if we consider the log-likelihood function we can rewrite δ_t as follows:

$$\delta_t = f(Y_t) \quad \text{where} \quad f(Y) := \frac{\delta}{\delta + (1-\delta)\,\exp(Y)}, \quad Y_t := \sum_{i=0}^{N_t} \beta_i \quad \text{and} \quad \beta_i := \ln\left(\mathcal{L}(x_{t_i}, \mathcal{E}_{t_i})\right).$$

Using Itô's lemma, we can express δ_t as the solution of the SDE

$$\begin{aligned}
\mathrm{d}\delta_t &= f'(Y_{t-})\,\mathrm{d}Y_t + f(Y_t) - f(Y_{t-}) - f'(Y_{t-})\,\Delta Y_t \\
&= f(Y_t) - f(Y_{t-}) \\
&= \big(f(Y_{t-} + \beta_{N_t}) - f(Y_{t-})\big)\,\mathrm{d}N_t \\
&= (1-\delta_{t-})\,\delta_{t-}\left(\frac{Q(x_t, \mathcal{E}_t, \theta_0) - Q(x_t, \mathcal{E}_t, \theta_1)}{Q(x_t, \mathcal{E}_t, \theta_0)\,\delta_{t-} + Q(x_t, \mathcal{E}_t, \theta_1)\,(1-\delta_{t-})}\right)\mathrm{d}N_t.
\end{aligned}$$

where the second equality follows from the fact that Y_t is a pure jump process, *i.e.*, $\mathrm{d}Y_t = \Delta Y_t$. □

Proof of Proposition 2: To prove the result we invoke Theorem 4.21 in Chapter IX in Jacod and Shiryaev (2003) related to the convergence of Markov processes to diffusions. To this end, note that from Lemma 2 it follows that the belief process δ_t^k is a pure jump Markov process that admits a generator of the form

$$\mathcal{A}^k f(\delta) = \int_{y\in(0,1)} \left[f(\delta + y) - f(\delta)\right] K^k(\delta, \mathrm{d}y),$$

where the kernel $K^k(\delta, y)$ satisfies

$$\int_{y\in(0,1)} f(y)\,K^k(\delta, \mathrm{d}y) = \Lambda^k \sum_{\mathcal{E}\in\mathscr{E}} \sum_{x\in\mathcal{E}} f(\eta^k(\delta, x, \mathcal{E}))\,Q_\delta^k(x, \mathcal{E})\,\pi(\delta, \mathcal{E})$$

where $Q^k_\delta(x,\mathcal{E}) := \delta\, Q^k(x,\mathcal{E},\theta_0) + (1-\delta)\, Q^k(x,\mathcal{E},\theta_1)$ and

$$\eta^k(\delta,x,\mathcal{E}) := (1-\delta)\,\delta\left(\frac{1-\mathcal{L}^k(x,\mathcal{E})}{\delta+(1-\delta)\,\mathcal{L}^k(x,\mathcal{E})}\right).$$

It follows that the instantaneous drift and volatility of δ^k_t are given by

$$b^k(\delta) := \int_{y\in(0,1)} y\, K^k(\delta,\mathrm{d}y) = \Lambda^k \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} \eta^k(\delta,x,\mathcal{E})\, Q^k_\delta(x,\mathcal{E})\,\pi(\delta,\mathcal{E}) = 0$$

and

$$\begin{aligned} c^k(\delta) := \int_{y\in(0,1)} y^2\, K^k(\delta,\mathrm{d}y) &= \Lambda^k \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} (\eta^k(\delta,x,\mathcal{E}))^2\, Q^k_\delta(x,\mathcal{E})\,\pi(\delta,\mathcal{E}) \\ &= \Lambda^k\,\delta\,(1-\delta) \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} \frac{\left(Q^k(x,\mathcal{E},\theta_0) - Q^k(x,\mathcal{E},\theta_1)\right)^2}{Q^k_\delta(x,\mathcal{E})}\,\pi(\delta,\mathcal{E}) \\ &= \Lambda\,\delta\,(1-\delta) \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} \frac{\left(\alpha^k(x,\mathcal{E},\theta_0) - \alpha^k(x,\mathcal{E},\theta_0)\right)^2 \mathcal{Q}^k(x,\mathcal{E})}{1+(\delta\,\alpha^k(x,\mathcal{E},\theta_0)+(1-\delta)\,\alpha^k(x,\mathcal{E},\theta_1))/\sqrt{k}}\,\pi(\delta,\mathcal{E}). \end{aligned}$$

It follows by Assumption 1 that

$$\begin{aligned} b(\delta) &:= \lim_{k\to\infty} b^k(\delta) = 0 \qquad \text{and} \\ c(\delta) &:= \lim_{k\to\infty} c^k(\delta) = \Lambda\,\delta\,(1-\delta) \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} \left(\alpha(x,\mathcal{E},\theta_0) - \alpha(x,\mathcal{E},\theta_1)\right)^2 \mathcal{Q}(x,\mathcal{E})\,\pi(\delta,\mathcal{E}). \end{aligned}$$

(Note that the convergence of $b^k(\delta)$ and $c^k(\delta)$ is trivially locally uniformly in $(0,1)$).
Since the jump size $\eta^k(\delta,x,\mathcal{E})$ converges to zero as $k\to\infty$ uniformly in δ for all $\mathcal{E}$ and all $x\in\mathcal{E}$, we get that for all $\epsilon>0$

$$\begin{aligned} \sup_{\delta\in(0,1)} \int_{y\in(0,1)} y^2\, \mathbb{1}(y>\epsilon)\, K^k(\delta,\mathrm{d}y) &= \Lambda^k \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}} (\eta^k(\delta,x,\mathcal{E}))^2\, \mathbb{1}(|\eta^k(\delta,x,\mathcal{E})|>\epsilon)\, Q^k_\delta(x,\mathcal{E})\,\pi(\delta,\mathcal{E}) \\ &\to 0 \qquad \text{as } k\to\infty. \end{aligned}$$

To conclude, note that $b(\delta)$ is trivially bounded and $c(\delta)$ is continuous and bounded in $(0,1)$ since are assuming that $\pi\in\mathcal{M}_c$. We also have $c(\delta)>0$ for all $\delta\in(0,1)$ since we have assumed that every experiment $\mathcal{E}$ is informative and so we must have $\sum_{x\in\mathcal{E}}\left(\alpha(x,\mathcal{E},\theta_0)-\alpha(x,\mathcal{E},\theta_1)\right)^2>0$. So, by Theorem 2.34 in Chapter III in Jacod and Shiryaev (2003) the semimartingale problem with characteristics (b,c) has a unique solution for every initial condition $\delta\in(0,1)$. In sum, all the required conditions in Theorem 4.21 in Chapter IX in Jacod and Shiryaev (2003) are satisfied and so δ^k_t converges weakly to a diffusion process $\tilde{\delta}_t$ with characteristics (b,c). □

Proof of Proposition 3: Consider an arbitrary instance of the problem, and let $(\pi,\mathcal{I})\in\mathcal{M}(\mathscr{E})\times\mathcal{B}$ be an optimal Markovian policy with corresponding value function $\Pi(\delta)$. For future references, we recall

that for any bounded function $f(\delta)$, the value-iteration recursion

$$\Pi_{j+1}(\delta) = \mathbb{1}(\delta \in \mathcal{I})\, G(\delta) + \mathbb{1}(\delta \in \mathcal{I}^c)\, \rho\, \mathbb{E}_\pi\Big[\Pi_j(\delta_1)|\delta_0 = \delta\Big], \quad \Pi_0(\delta) = f(\delta), \text{ where } \rho := \frac{\Lambda}{\Lambda + r} \tag{A-2}$$

produces a sequence of functions $\{\Pi_j\}_{j\geq 0}$ that converges pointwise to Π. In (A-2), δ_1 denotes the value of the belief process after one jump (vote) and $\mathbb{E}_\pi[\cdot]$ is the expectation operator induced by policy $(\pi, \mathcal{I})$, which satisfies

$$\mathbb{E}_\pi[f(\delta_{j+1})|\delta_j = \delta] = \mathbb{1}(\delta \in \mathcal{I})\, f(\delta) + \mathbb{1}(\delta \in \mathcal{I}^c) \sum_{\mathcal{E}\in\mathscr{E}} \sum_{x\in\mathcal{E}} f(\delta + \eta(\delta, x, \mathcal{E}))\, Q_\delta(x, \mathcal{E})\, \pi(\delta, \mathcal{E}).$$

Note that in (A-2) and in the definition of $\mathbb{E}_\pi[\cdot]$ have used the fact that the set $\mathcal{I}$ is absorbing under policy $(\pi, \mathcal{I})$. Also, the fact that $\{\Pi_j\}_{j\geq 0}$ converges pointwise to Π follows by a standard contraction mapping argument (e.g., Chapter 6 in Puterman, 2005).

Now, since the class of continuous experimentation strategies $\mathcal{M}_c(\Delta\mathscr{E})$ is dense in $\mathcal{M}(\Delta\mathscr{E})$ under the L^1 norm, there exists a sequence of continuous strategies $\{\hat{\pi}_n\}_{n\geq 1}$ in $\mathcal{M}_c(\Delta\mathscr{E})$ that converges in L^1 to π. Let us denote by $\widehat{\Pi}_n$ be the expected payoff function under the policy $(\hat{\pi}_n, \mathcal{I})$. It follows that for each n, the function $\widehat{\Pi}_n$ satisfies the fixed-point condition

$$\widehat{\Pi}_n(\delta) = \mathbb{1}(\delta \in \mathcal{I})\, G(\delta) + \mathbb{1}(\delta \in \mathcal{I}^c)\, \rho\, \mathbb{E}_{\hat{\pi}_n}\Big[\widehat{\Pi}_n(\delta_1)|\delta_0 = \delta\Big], \tag{A-3}$$

where the expectation operator $\mathbb{E}_{\hat{\pi}_n}[\cdot]$ under policy $(\hat{\pi}_n, \mathcal{I})$ is defined in a similar way to $\mathbb{E}_\pi[\cdot]$ above. One fact to keep in mind is that by the optimality of $(\pi, \mathcal{I})$ we have that $\Pi(\delta) \geq \widehat{\Pi}_n(\delta)$.

We want to show that $\widehat{\Pi}_n$ converges to $\Pi(\delta)$ in L^1 as $n \to \infty$. To this end, let us use the recursion in (A-2) with initial condition $\Pi_0(\delta) = \widehat{\Pi}_n(\delta)$. Combining (A-2) and (A-3), one can show that

$$\Pi_1(\delta) = \widehat{\Pi}_n(\delta) + \mathbb{1}(\delta \in \mathcal{I}^c)\, \rho \sum_{\mathcal{E}\in\mathscr{E}} \sum_{x\in\mathcal{E}} \widehat{\Pi}_n(\delta + \eta(\delta, x, \mathcal{E}))\, Q_\delta(x, \mathcal{E})\, [\pi(\delta, \mathcal{E}) - \hat{\pi}_n(\delta, \mathcal{E})] \leq \widehat{\Pi}_n(\delta) + F_n(\delta),$$

where

$$F_n(\delta) := \mathbb{1}(\delta \in \mathcal{I}^c)\, \rho \max_\delta \Big\{\widehat{\Pi}_n(\delta)\Big\} \sum_{\mathcal{E}\in\mathscr{E}} \big|\pi(\delta, \mathcal{E}) - \hat{\pi}_n(\delta, \mathcal{E})\big|.$$

We note that $\|F_n\|_1 \leq K\|\pi - \hat{\pi}_n\|_1$ for some fixed constant K. (The fact that we can choose K independent of n follows from the fact that the $\widehat{\Pi}_n$ are uniformly bounded above by Π.)

Let us iterate the recursion (A-2) one more time for Π_2. Using the inequality $\Pi_1(\delta) \leq \widehat{\Pi}_n(\delta) + F_n(\delta)$, we get that

$$\Pi_2(\delta) \leq \widehat{\Pi}_n(\delta) + F_n(\delta) + \rho\, \mathbb{E}_\pi\Big[F_n(\delta_1)|\delta_0 = \delta\Big].$$

If we keep iterating this inequality we get that

$$\Pi_j(\delta) \leq \widehat{\Pi}_n(\delta) + \mathbb{E}_\pi\left[\sum_{\ell=0}^{j-1} \rho^\ell\, F_n(\delta_\ell)\big|\delta_0 = \delta\right],$$

where δ_ℓ denotes the state of the belief process after ℓ jumps (votes). Let us denote by J the random

time at which δ_ℓ enters $\mathcal{I}$, that is, $J := \inf\{\ell \geq 0 \colon \delta_\ell \in \mathcal{I}\}$. Since $F_n(\delta_\ell) = 0$ for all $\ell \geq J$ and $F_n(\delta) \geq 0$, we have that

$$\Pi_j(\delta) \leq \widehat{\Pi}_n(\delta) + \mathbb{E}_\pi \left[\sum_{\ell=0}^{J-1} \rho^\ell \, F_n(\delta_\ell) \big| \delta_0 = \delta \right].$$

Hence, taking limit as $j \uparrow \infty$ and using the pointwise convergence of $\{\Pi_j\}$ to Π, we get

$$\Pi(\delta) \leq \widehat{\Pi}_n(\delta) + \mathbb{E}_\pi \left[\sum_{\ell=0}^{J-1} \rho^\ell \, F_n(\delta_\ell) \big| \delta_0 = \delta \right].$$

Since Π is the optimal value function, it follows that $\Pi(\delta) \geq \widehat{\Pi}_n(\delta)$ and so

$$\|\Pi - \widehat{\Pi}_n\|_1 \leq \int_0^1 \mathbb{E}_\pi \left[\sum_{\ell=0}^{J-1} \rho^\ell \, F_n(\delta_\ell) \big| \delta_0 = \delta \right] \mathrm{d}\delta.$$

Next, we use a localization argument. Let us define $\bar{F} := \sup_n \max_\delta \{F_n(\delta)\}$ and let T be a fixed nonnegative integer. Then, the previous inequality implies

$$\|\Pi - \widehat{\Pi}_n\|_1 \leq \sum_{\ell=0}^{T-1} \rho^\ell \int_0^1 \mathbb{E}_\pi[F_n(\delta_\ell) \big| \delta_0 = \delta] \, \mathrm{d}\delta + \bar{F} \frac{\Lambda}{r} \, \rho^{T-1} \, \mathbb{E}_\pi \left[1 - \rho^{(J-T)^+} \right].$$

To complete the proof, we will show in Lemma 3 below that there exists a constant $\bar{\mu} > 0$ such that

$$\int_0^1 \mathbb{E}_\pi[F_n(\delta_\ell) \big| \delta_0 = \delta] \, \mathrm{d}\delta \leq \bar{\mu}^\ell \, \|F_n\|_1 \leq K \, \bar{\mu}^\ell \, \|\pi - \hat{\pi}_n\|_1.$$

As a result,

$$\|\Pi - \widehat{\Pi}_n\|_1 \leq K \left(\frac{1 - (\rho \, \bar{\mu})^T}{1 - \rho \, \bar{\mu}} \right) \|\pi - \hat{\pi}_n\|_1 + \bar{F} \frac{\Lambda}{r} \, \rho^{T-1} \, \mathbb{E}_\pi \left[1 - \rho^{(J-T)^+} \right].$$

Taking limit as $n \uparrow \infty$ and using the fact that $\|\pi - \hat{\pi}_n\|_1 \downarrow 0$ as $n \uparrow \infty$ we get that

$$\lim_{n \to \infty} \|\Pi - \widehat{\Pi}_n\|_1 \leq \bar{F} \frac{\Lambda}{r} \, \rho^{T-1} \, \mathbb{E}_\pi \left[1 - \rho^{(J-T)^+} \right].$$

Finally, since T is arbitrary, we can let $T \uparrow \infty$ to complete the proof. □

Lemma 3. *There exists a constant $\bar{\mu} > 0$ such that for any non-negative function $f(\delta)$ with $f(\delta) = 0$ in $\mathcal{I}$*

$$\int_0^1 \mathbb{E}_\pi[f(\delta_\ell) \big| \delta_0 = \delta] \, \mathrm{d}\delta \leq \bar{\mu}^\ell \, \|f\|_1.$$

Proof of Lemma 3: We use a proof by induction. Let us consider first the case $\ell = 1$. From the definition of

$\mathbb{E}_\pi[\cdot]$ we have that

$$\begin{aligned}
\int_0^1 \mathbb{E}_\pi[f(\delta_\ell)\big|\delta_0=\delta]\,\mathrm{d}\delta &\leq \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}}\int_0^1 f(\delta+\eta(\delta,x,\mathcal{E}))\,Q_\delta(x,\mathcal{E})\,\pi(\delta,\mathcal{E})\,\mathrm{d}\delta \\
&= \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}}\int_0^1 f\left(\frac{\delta}{\delta+(1-\delta)\,\mathcal{L}(x,\mathcal{E})}\right)\,Q_\delta(x,\mathcal{E})\,\pi(\delta,\mathcal{E})\,\mathrm{d}\delta \\
&= \sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}}\int_0^1 \frac{f\,(u)\;Q_{\delta_u}(x,\mathcal{E})\,\pi(\delta_u,\mathcal{E})\,\mathcal{L}(x,\mathcal{E})}{(1+\mathcal{L}(x,\mathcal{E})\,u-u)^2}\mathrm{d}u \quad \text{with } \delta_u=\frac{\mathcal{L}(x,\mathcal{E})\,u}{1+\mathcal{L}(x,\mathcal{E})\,u-u} \\
&\leq \max_{\mathcal{E}\in\mathscr{E}}\max_{x\in\mathcal{E}}\left\{\mathcal{L}(x,\mathcal{E}),\frac{1}{\mathcal{L}(x,\mathcal{E})}\right\}\sum_{\mathcal{E}\in\mathscr{E}}\sum_{x\in\mathcal{E}}\int_0^1 f\,(u)\;Q_{\delta_u}(x,\mathcal{E})\,\pi(\delta_u,\mathcal{E})\mathrm{d}u \\
&= \max_{\mathcal{E}\in\mathscr{E}}\max_{x\in\mathcal{E}}\left\{\mathcal{L}(x,\mathcal{E}),\frac{1}{\mathcal{L}(x,\mathcal{E})}\right\}\int_0^1 f(u)\,\mathrm{d}u. \qquad \text{(A-4)}
\end{aligned}$$

In the second inequality we have used the fact that

$$\max_{u\in[0,1]}\frac{L}{(1+L\,u-u)^2}\leq\max\left\{L,\frac{1}{L}\right\}.$$

So, the result in Lemma 3 holds for $\ell=1$ with

$$\bar{\mu}:=\max_{\mathcal{E}\in\mathscr{E}}\max_{x\in\mathcal{E}}\left\{\mathcal{L}(x,\mathcal{E}),\frac{1}{\mathcal{L}(x,\mathcal{E})}\right\}.$$

Suppose that the result in Lemma 3 is true for $j=1,\ldots,\ell-1$. From the law of iterated expectations we have that $\mathbb{E}_\pi[f(\delta_\ell)|\delta_0=\delta]=\mathbb{E}_\pi[\mathbb{E}_\pi[f(\delta_\ell)|\delta_{\ell-1}]|\delta_0=\delta]=\mathbb{E}_\pi[g(\delta_{\ell-1})|\delta_0=\delta]$ with $g(\delta):=\mathbb{E}_\pi[f(\delta_\ell)|\delta_{\ell-1}=\delta]$. But, by the Markov property this is the same as $g(\delta):=\mathbb{E}_\pi[f(\delta_1)|\delta_0=\delta]$. It follows from the hypothesis of induction and the inequality (A-4) that

$$\begin{aligned}
\int_0^1 \mathbb{E}_\pi[f(\delta_\ell)\big|\delta_0=\delta]\,\mathrm{d}\delta &= \int_0^1 \mathbb{E}_\pi[g(\delta_{\ell-1})|\delta_0=\delta]\,\mathrm{d}\delta \leq \bar{\mu}^{\ell-1}\int_0^1 g(\delta)\,\mathrm{d}\delta=\bar{\mu}^{\ell-1}\int_0^1 \mathbb{E}_\pi[f(\delta_1)|\delta_0=\delta]\,\mathrm{d}\delta \\
&\leq \bar{\mu}^\ell\,\|f\|_1. \quad \square
\end{aligned}$$

Proof of Proposition 4: For a given experimentation policy π consider the mapping

$$T_t^\pi:=\int_0^t\frac{1}{\tilde{\sigma}^2(\tilde{\delta}_s,\pi)}\,\mathrm{d}s,$$

where $\tilde{\sigma}^2(\delta,\pi)$ is defined in equation (10). Since, $\tilde{\sigma}^2(\delta,\pi)>0$ for all δ, the mapping T_t^π is strictly increasing t with $T_0^\pi=0$. As a result, let us view T_t^π as a random time change and let us define the process

$$\hat{\delta}_t:=\tilde{\delta}_{T_t^\pi}.$$

Let $\mathcal{G}_{\tilde{\delta}}$ denote the infinitesimal generator of $\tilde{\delta}_t$. Then, by Proposition 2 it follows that

$$\mathcal{G}_{\tilde{\delta}} = \tilde{\sigma}^2(\delta, \pi)\, \delta^2\, (1-\delta)^2\, \frac{\partial^2}{\partial \delta^2}.$$

It follows that the infinitesimal generator $\mathcal{G}_{\hat{\delta}}$ of $\hat{\delta}_t$ is given by

$$\mathcal{G}_{\hat{\delta}} = \dot{T}_t^{\pi}\, \mathcal{G}_{\tilde{\delta}} = \frac{1}{\tilde{\sigma}^2(\delta, \pi)}\, \tilde{\sigma}^2(\delta, \pi)\, \delta^2\, (1-\delta)^2\, \frac{\partial^2}{\partial \delta^2} = \delta^2\, (1-\delta)^2\, \frac{\partial^2}{\partial \delta^2}.$$

In other words, $\hat{\delta}_t$ is a diffusion process that satisfies the SDE

$$\mathrm{d}\hat{\delta}_t = \hat{\delta}_t\, (1-\hat{\delta})\, \mathrm{d}W_t, \tag{A-5}$$

for some Wiener process W_t. Also, for a given stopping time τ for $\tilde{\delta}_t$, let us define the stopping time $\hat{\tau}$ for $\hat{\delta}_t$ such that $\hat{\delta}_{\hat{\tau}} = \tilde{\delta}_{\tau}$. It follows that

$$\tau = \int_0^{\hat{\tau}} \frac{1}{\tilde{\sigma}^2(\tilde{\delta}_s, \pi)}\, \mathrm{d}s. \tag{A-6}$$

Finally, the result in Proposition 4 follows from equations (A-5) and (A-6). □

Proof of Theorem 1: Let f be a solution to the QVI in equation (15). Given the assumptions on f, we can apply integration by parts followed by Itô's lemma (see Protter, 2004) to get that

$$e^{-r\,\tau}\, f(\delta_\tau) = f(\delta) + \int_0^{\tau} e^{-r\,t}\, \mathcal{H} f(\delta_t)\, \mathrm{d}t + \int_0^{\tau} e^{-r\,t} \tilde{\sigma}\, \delta_t\, (1-\delta_t)\, f'(\delta_t)\, \mathrm{d}W_t.$$

Note that the process

$$f(\delta) + \int_0^{t} e^{-r\,s} \tilde{\sigma}\, \delta_s\, (1-\delta_s)\, f'(\delta_s)\, \mathrm{d}W_s,$$

is a local martingale, thus, by the non-negativity of f, also a supermartingale. With this, one can take expectation, canceling the stochastic integral, and use the fact that $\mathcal{H} f(\delta) \leq 0$ (second QVI condition) to get that

$$\mathbb{E}[e^{-r\,\tau}\, f(\delta_\tau)] \leq f(\delta).$$

This inequality together with the first QVI condition imply

$$\mathbb{E}[e^{-r\,\tau}\, \widetilde{G}(\delta_\tau)] \leq \mathbb{E}[e^{-r\,\tau}\, f(\delta_\tau)] \leq f(\delta).$$

Because these inequalities hold for any stopping time τ, we conclude that $f(\delta) \geq \widetilde{\mathcal{G}}(\delta)$. Finally, we note that all the inequalities above become equalities for the QVI-control associated to f. This follows from Dynkin's formula and the fact that the QVI-control is the first exit time from a bounded set (continuation region $\mathcal{C}$). □

Proof of Proposition 5: We need to prove both the existence and optimality of the function $\widetilde{\mathcal{G}}_{ij}(\delta)$

in equation (18). Let us start by proving the optimality using the QVI conditions.

The first step is to show that $\widetilde{\mathcal{G}}_{ij}(\delta)$ is convex in $[0,1]$. To see this note that in the continuation region $\delta \in (\underline{\delta}_{ij}, \bar{\delta}_{ij})$ the function $\widetilde{\mathcal{G}}_{ij}(\delta) = C^0_{ij}\,(1-\delta)^{\gamma}\,\delta^{1-\gamma} + C^1_{ij}\,(1-\delta)^{1-\gamma}\,\delta^{\gamma}$ is convex. This follows from the fact that by construction it satisfies the ODE $\mathcal{H}\widetilde{\mathcal{G}}_{ij}(\delta) = 0$, and so

$$\frac{(\tilde{\sigma}\,\delta\,(1-\delta))^2}{2}\,\widetilde{\mathcal{G}}''_{ij}(\delta) = r\,\widetilde{\mathcal{G}}_{ij}(\delta) \geq 0$$

This together with the value matching and smooth pasting conditions at $\bar{\delta}_{ij}$ and $\bar{\delta}_{ij}$ ensure that $\widetilde{\mathcal{G}}_{ij}(\delta)$ is convex in $[0,1]$.

Now, by convexity and the smooth-pasting and value matching conditions, both $\widetilde{\mathcal{R}}_i(\delta)$ and $\widetilde{\mathcal{R}}_j(\delta)$ are supporting hyperplanes of $\widetilde{\mathcal{G}}_{ij}(\delta)$ in the domain $\delta \in [0,1]$. We conclude that the first QVI condition holds, that is, $\widetilde{\mathcal{G}}_{ij}(\delta) \geq \widetilde{G}_{ij}(\delta) = \max\{\widetilde{\mathcal{R}}_i(\delta), \widetilde{\mathcal{R}}_j(\delta)\}$.

To prove the second and third QVI conditions note that in the continuation region $\delta \in (\underline{\delta}_{ij}, \bar{\delta}_{ij})$, we have $\mathcal{H}\widetilde{\mathcal{G}}_{ij}(\delta) = 0$ (by construction). On the other hand, in the intervention region $\delta \in [0, \underline{\delta}_{ij}] \cup [\bar{\delta}_{ij}, 1]$, we have that (a) $\widetilde{\mathcal{G}}_{ij}(\delta) = \widetilde{G}_{ij}(\delta)$ and (b) $\mathcal{H}\widetilde{\mathcal{G}}_{ij}(\delta) = -r\,\widetilde{G}_{ij}(\delta) \leq 0$.

Finally, if we define the set $N_{ij} = \{\underline{\delta}_{ij}, \bar{\delta}_{ij}, \hat{\delta}_{ij}\}$, it is easy to see that the function $\widetilde{\mathcal{G}}_{ij}(\delta)$ is in $\mathcal{C}^1[0,1]$ and has second derivative for all $\delta \in [0,1] \setminus N_{ij}$. We conclude that $\widetilde{\mathcal{G}}_{ij}(\delta) \in \hat{\mathcal{C}}^2$ and so by Theorem 1 it is optimal.

Let us now turn to the issue of existence. For this, we need to show that there exist thresholds $\underline{\delta}_{ij}$ and $\bar{\delta}_{ij}$ so that the smooth pasting and value matching conditions are satisfied. To fix ideas, let us suppose that $\widetilde{\mathcal{R}}_i(0) \geq \widetilde{\mathcal{R}}_j(0)$ and let us consider the auxiliary function

$$V(\delta;\underline{\delta}) := \begin{cases} \widetilde{\mathcal{R}}_i(\delta) & \text{if} \quad 0 \leq \delta \leq \underline{\delta} \\ \\ C^0_{ij}(\underline{\delta})\,(1-\delta)^{\gamma}\,\delta^{1-\gamma} + C^1_{ij}(\underline{\delta})\,(1-\delta)^{1-\gamma}\,\delta^{\gamma} & \text{if} \quad \underline{\delta} \leq \delta \leq 1 \end{cases}$$

where the parameter $\underline{\delta} \in [0, \hat{\delta}_{ij}]$ and the constants $C^0_{ij}(\underline{\delta})$ and $C^1_{ij}(\underline{\delta})$ are chosen to ensure value matching and smooth pasting at $\delta = \underline{\delta}$. Recall that the payoff associated to each action is linear in δ, that is, of the form $\widetilde{\mathcal{R}}_i(\delta) = \tilde{\alpha}_i + \tilde{\beta}_i\,\delta$. It follows that the constants $C^0_{ij}(\underline{\delta})$ and $C^1_{ij}(\underline{\delta})$ are equal to

$$C^0_{ij}(\underline{\delta}) = \left[\frac{(\gamma-1)\,\underline{\delta}\,\tilde{\beta}_i + (\gamma-\underline{\delta})\,\tilde{\alpha}_i}{(2\,\gamma-1)\,\underline{\delta}}\right]\left(\frac{\underline{\delta}}{1-\underline{\delta}}\right)^{\gamma} \quad \text{and} \quad C^1_{ij}(\underline{\delta}) = \left[\frac{\gamma\,\underline{\delta}\,\tilde{\beta}_i + (\gamma-1+\underline{\delta})\,\tilde{\alpha}_i}{(2\,\gamma-1)\,\underline{\delta}}\right]\left(\frac{1-\underline{\delta}}{\underline{\delta}}\right)^{\gamma-1}.$$

Of course, by construction the function $V(\delta;\underline{\delta})$ satisfies the value matching and smooth pasting conditions at $\underline{\delta}$. Next, we show that by varying the value of $\underline{\delta}$ we can also enforce these conditions at the upper threshold $\bar{\delta}$. To get some intuition, consider the example in Figure 6 which depicts the function $V(\delta;\underline{\delta})$ for three different values of $\underline{\delta} \in \{0.1, 0.295, 0.43\}$. The figure also shows the payoff functions $\widetilde{\mathcal{R}}_i(\delta)$ and $\widetilde{\mathcal{R}}_j(\delta)$. We note that when $\underline{\delta}$ is small (in the example $\underline{\delta} = 0.1$) the function $V(\delta;\underline{\delta})$ is greater than $\widetilde{\mathcal{R}}_j(\delta)$ for all $\delta \geq \underline{\delta}$. On the opposite case, when $\underline{\delta}$ is large (in the example $\underline{\delta} = 0.43$) the function $V(\delta;\underline{\delta})$ intersects $\widetilde{\mathcal{R}}_j(\delta)$ for some $\delta \geq \underline{\delta}$. By continuity, there is a value of $\underline{\delta}$ (in the example $\underline{\delta} = 0.295$) so that $V(\delta;\underline{\delta})$ and $\widetilde{\mathcal{R}}_j(\delta)$ meet smoothly at some $\bar{\delta} \geq \underline{\delta}$.

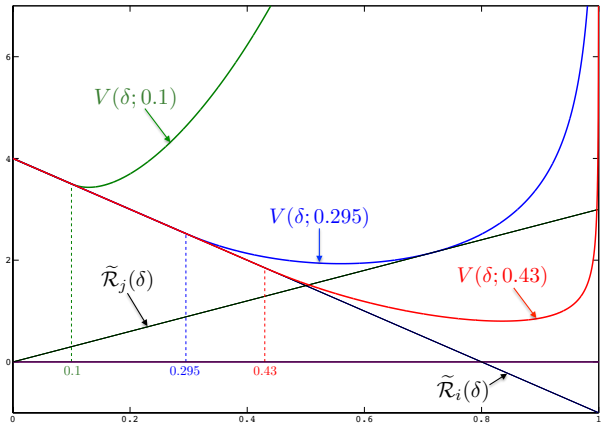

Figure 6: Value of $V(\delta;\underline{\delta})$ for three values of $\underline{\delta} \in \{0.1, 0.295, 0.43\}$. For $\underline{\delta} = 0.295$, the function $V(\delta;\underline{\delta})$ satisfies the smooth-pasting condition at $\bar{\delta}$. DATA: $\widetilde{\mathcal{R}}_j(\delta) = 3\,\delta$, $\widetilde{\mathcal{R}}_i(\delta) = 4 - 5\,\delta$, $r = 1$ and $\tilde{\sigma} = 2$.

To formalize the previous discussion based on the example in Figure 6, let us first note that $\lim_{\underline{\delta}\downarrow 0} C^0_{ij}(\underline{\delta}) = 0$ and $\lim_{\underline{\delta}\downarrow 0} C^1_{ij}(\underline{\delta}) = \infty$ (recall that $\gamma > 1$). Hence, the $\lim_{\underline{\delta}\downarrow 0} V(\delta,\underline{\delta}) = \infty$ for all $\delta \in (0,1]$. This shows that if $\underline{\delta}$ is sufficiently small the function $V(\delta,\underline{\delta})$ will be strictly greater than $\widetilde{\mathcal{R}}_j(\delta)$ for all $\delta \geq \underline{\delta}$.
On the flip side, we have that $\lim_{\underline{\delta}\to\hat{\delta}_{ij}} V(\underline{\delta},\underline{\delta}) = \widetilde{\mathcal{R}}_j(\underline{\delta})$ and $\lim_{\underline{\delta}\to\hat{\delta}_{ij}} V'(\underline{\delta},\underline{\delta}) = \tilde{\beta}_i < \tilde{\beta}_j = \widetilde{\mathcal{R}}'_j(\underline{\delta})$. In other words, if $\underline{\delta}$ is sufficiently close to $\hat{\delta}_{ij}$ then the function $V(\delta,\underline{\delta})$ will intersect and go below the function $\widetilde{\mathcal{R}}_j(\delta)$.
Finally, since the function $V(\delta,\underline{\delta})$ is continues in $\underline{\delta}$, we conclude that there exists a value $\underline{\delta} = \underline{\delta}_{ij} \in [0, \hat{\delta}_{ij}]$ such that

$$\min_{\delta\in[\hat{\delta}_{ij},1]} \left\{V(\delta,\underline{\delta}_{ij}) - \widetilde{\mathcal{R}}_j(\delta)\right\} = 0.$$

The value of δ that solves the minimization is the upper threshold $\bar{\delta}_{ij}$. □

PROOF OF COROLLARY 2: For notational convenience, let us write $\underline{\delta} = \underline{\delta}_{ij}$ and $\bar{\delta} = \bar{\delta}_{ij}$.
To compute the probability $\bar{p}(\delta) = \mathbb{P}(\tilde{\delta}_{\tau^*} = \bar{\delta} | \tilde{\delta}_0 = \delta)$, we use Dynkin's formula to get

$$\mathbb{E}[f(\tilde{\delta}_{\tau^*})] = f(\delta) + \mathbb{E}\left[\int_0^{\tau^*} \mathcal{G}f(\tilde{\delta}_t)\,\mathrm{d}t\right], \qquad \text{where} \quad \mathcal{G}f(\delta) := \frac{(\tilde{\sigma}\,\delta\,(1-\delta))^2}{2}\,\frac{\mathrm{d}^2 f(\delta)}{\mathrm{d}\delta^2},$$

and $\mathcal{G}$ is the infinitesimal generator of the diffusion process $\tilde{\delta}_t$ in equation (12). Consider the identity function $f(\delta) = \delta$. It follows that $\mathcal{G}f(\delta) = 0$ and by Dynkin's formula $\mathbb{E}[\tilde{\delta}_{\tau^*}] = \delta$. But since τ^* is the first exit time of the process $\tilde{\delta}_t$ from the continuation region $(\underline{\delta},\bar{\delta})$ we have that $\mathbb{E}[\tilde{\delta}_{\tau^*}] = \bar{p}(\delta)\,\bar{\delta} + (1 - \bar{p}(\delta))\,\underline{\delta}$ and the result part of the Corollary follows.
To compute the expectation $\mathbb{E}[\tau^*]$, we consider a function $\mathcal{T}(\delta)$ such that $\mathcal{G}(\mathcal{T})(\delta) = 1$. One can verify that the function $\mathcal{T}(\delta) = \frac{2}{\tilde{\sigma}^2}\,(2\delta - 1)\,\ln\left(\frac{\delta}{1-\delta}\right)$ satisfies this condition. It follows from Dynkin's formula that $\mathbb{E}[\mathcal{T}(\tilde{\delta}_{\tau^*})] = \mathcal{T}(\delta) + \mathbb{E}[\tau^*]$ and the result follows. □

PROOF OF PROPOSITION 7: By Proposition 5, if follows that $\widetilde{\mathcal{G}}_{ij}(\delta) \geq \widetilde{G}_{ij}(\delta)$ for all $\delta \in [0,1]$. As a result, $\widetilde{V}(\delta) = \max_{\{i,j\}\in\widetilde{\mathcal{O}}}\{\widetilde{\mathcal{G}}_{ij}(\delta)\} \geq \max_{\{i,j\}\in\widetilde{\mathcal{O}}}\{\widetilde{G}_{ij}(\delta)\} = \widetilde{G}(\delta)$.

By Proposition 5, we know that each function $\widetilde{\mathcal{G}}_{ij}(\delta)$ is continuously differentiable everywhere in $[0,1]$ and admits a second derivative almost everywhere in $[0,1]$ except in the set $N_{ij} := \{\underline{\delta}_{ij}, \bar{\delta}_{ij}\}$. Also, two functions $\widetilde{\mathcal{G}}_{ij}(\delta)$ and $\widetilde{\mathcal{G}}_{k\ell}(\delta)$ can cross at most a finite number of times. This follows from noticing that in the continuation regions the functions $C^0_{ij}\,(1-\delta)^\gamma\,\delta^{1-\gamma} + C^1_{ij}\,(1-\delta)^{1-\gamma}\,\delta^\gamma$ and $C^0_{k\ell}\,(1-\delta)^\gamma\,\delta^{1-\gamma} + C^1_{k\ell}\,(1-\delta)^{1-\gamma}\,\delta^\gamma$ can only cross at most once. Let us denote by $E_{ij,k\ell}$ the finite set of values of δ at which these two functions cross (if any) and let us define

$$N_{\widetilde{V}} = \bigcup_{i,j\in\widetilde{\mathcal{O}}} N_{ij} \cup \bigcup_{i,j,k,\ell\in\widetilde{\mathcal{O}}} \{E_{ij,k\ell}\}.$$

Now, by our previous construction, it follows that for each $\delta \in [0,1]\backslash N_{\widetilde{V}}$ there exists an open neighborhood $B(\delta)$ containing δ such that $\widetilde{V}(x) = \widetilde{\mathcal{G}}_{ij}(x)$ for all $x \in B(\delta)$ for some pair $\{i,j\} \in \widetilde{\mathcal{O}}$. Furthermore, the function $\widetilde{\mathcal{G}}_{ij}(x)$ is twice-continuously differentiable in $B(\delta)$. Since each function $\widetilde{\mathcal{G}}_{ij}(x)$ satisfies the QVI conditions, we conclude that $\mathcal{H}\widetilde{V}(\delta) \leq 0$ and $\big(\widetilde{V}(\delta) - \widetilde{G}(\delta)\big)\,\mathcal{H}\widetilde{V}(\delta) = 0$ for all $\delta \in [0,1] \setminus N_{\widetilde{V}}$. □

PROOF OF THEOREM 2: We wish to prove that the function $\widetilde{V}(\delta) = \max_{\{i,j\}\in\widetilde{\mathcal{O}}}\left\{\widetilde{\mathcal{G}}_{ij}(\delta)\right\}$ is in $\widehat{\mathcal{C}}^2$, where each function $\widetilde{\mathcal{G}}_{ij}(\delta)$ is convex and in $\widehat{\mathcal{C}}^2$ (Proposition 5). Furthermore, each function $\widetilde{\mathcal{G}}_{ij}(\delta)$ solves the optimization problem

$$\widetilde{\mathcal{G}}_{ij}(\delta) = \sup_{\tau\in\mathbb{T}} \mathbb{E}\left[e^{-r\,\tau}\,\max\left\{\widetilde{\mathcal{R}}_i(\delta), \widetilde{\mathcal{R}}_j(\delta)\right\}\right] \qquad \text{subject to} \qquad \mathrm{d}\,\tilde{\delta}_t = \tilde{\sigma}\,\tilde{\delta}_t\,(1-\tilde{\delta}_t)\,\mathrm{d}W_t, \qquad \tilde{\delta}_0 = \delta. \quad \text{(A-7)}$$

Let us suppose, by contradiction, that $\widetilde{V}(\delta)$ is not in $\widehat{\mathcal{C}}^2$, then it must exist a $\delta_\star$ at which $\widetilde{V}(\delta)$ is not differentiable. But since each $\widetilde{\mathcal{G}}_{ij}(\delta)$ is smooth it follows that there are at least two functions $\widetilde{\mathcal{G}}_{ij}(\delta)$ and $\widetilde{\mathcal{G}}_{k\ell}(\delta)$ that intersect and that simultaneously solve the maximization in the definition of $\widetilde{V}(\delta)$ at this value $\delta_\star$. That is,

$$\widetilde{V}(\delta_\star) = \widetilde{\mathcal{G}}_{ij}(\delta_\star) = \widetilde{\mathcal{G}}_{k\ell}(\delta_\star) \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}\delta}\widetilde{\mathcal{G}}_{ij}(\delta_\star) \neq \frac{\mathrm{d}}{\mathrm{d}\delta}\widetilde{\mathcal{G}}_{k\ell}(\delta_\star).$$

To fix ideas, let us suppose that $\widetilde{V}(\delta) = \widetilde{\mathcal{G}}_{ij}(\delta)$ for all $\delta \in (\delta_\star - \epsilon, \delta_\star]$ and $\widetilde{V}(\delta) = \widetilde{\mathcal{G}}_{k\ell}(\delta)$ for all $\delta \in [\delta_\star, \delta_\star + \epsilon)$ for some small $\epsilon > 0$ (as in Figure 7). In this case, the two conditions above imply that $\widetilde{\mathcal{G}}_{k\ell}(\delta_\star + \epsilon) > \widetilde{\mathcal{G}}_{ij}(\delta_\star + \epsilon)$ (i.e., point B is above point C).
We will now show that point D cannot belong to $\widetilde{V}(\delta)$. For this, we will exploit the optimality of the functions $\widetilde{\mathcal{G}}_{ij}(\delta)$ and $\widetilde{\mathcal{G}}_{k\ell}(\delta)$ in the sense of equation (A-7). We distinguish two cases:

1. Suppose that $\delta_\star$ belongs to at least one of the continuation regions $\mathcal{C}_{ij} = (\underline{\delta}_{ij}, \bar{\delta}_{ij})$ or $\mathcal{C}_{k\ell} = (\underline{\delta}_{k\ell}, \bar{\delta}_{k\ell})$ associated to $\widetilde{\mathcal{G}}_{ij}(\delta)$ and $\widetilde{\mathcal{G}}_{k\ell}(\delta)$, respectively (see equation (18) in Proposition 5). For concreteness let us assume that $\delta_\star \in \mathcal{C}_{ij}$.

   By choosing ϵ small enough we can guarantee that both $\delta_\star - \epsilon$ and $\delta_\star + \epsilon$ also belong to $\mathcal{C}_{ij}$. It

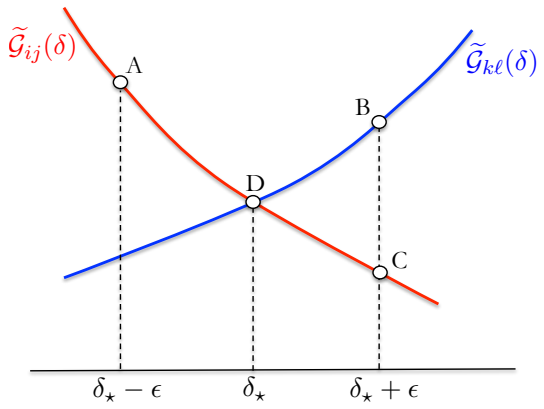

Figure 7: Schematic of what needs to happen for $\widetilde{V}(\delta) = \max\left\{\widetilde{\mathcal{G}}_{ij}(\delta), \widetilde{\mathcal{G}}_{k\ell}(\delta)\right\}$ to be non-smooth at some point $\delta_\star$.

follows then, by the principle of optimality, that

$$\widetilde{\mathcal{G}}_{ij}(\delta_\star) = \mathbb{E}\left[e^{-r\,\tau_\star}\,\widetilde{\mathcal{G}}_{ij}(\tilde{\delta}_{\tau_\star})\right], \quad \text{where} \quad \tau_\star := \inf\left\{t > 0 : \tilde{\delta}_t \notin (\delta_\star - \epsilon, \delta_\star + \epsilon)\right\}.$$

In words, this identity states that the value $\widetilde{\mathcal{G}}_{ij}(\delta_\star)$ can be obtained by letting the belief process $\tilde{\delta}_t$ evolve in the region $(\delta_\star - \epsilon, \delta_\star + \epsilon)$ and as soon as one of the boundaries is hit, then the corresponding value of $\widetilde{\mathcal{G}}_{ij}(\delta)$ is collected (point A if the left boundary is hit first or point C if the right boundary is hit first).

Now, using the fact that $\widetilde{\mathcal{G}}_{k\ell}(\delta_\star + \epsilon) > \widetilde{\mathcal{G}}_{ij}(\delta_\star + \epsilon)$ we get that

$$\widetilde{\mathcal{G}}_{ij}(\delta_\star) < \mathbb{E}\left[e^{-r\,\tau_\star}\,\max\left\{\widetilde{\mathcal{G}}_{ij}(\tilde{\delta}_{\tau_\star}), \widetilde{\mathcal{G}}_{k\ell}(\tilde{\delta}_{\tau_\star})\right\}\right].$$

But $\widetilde{V}(\delta_\star) = \widetilde{\mathcal{G}}_{ij}(\delta_\star)$ and so the previous inequality contradicts the optimality of $\widetilde{V}(\delta_\star)$.

2. Let us now suppose that $\delta_\star$ belong to both intervention regions $\mathcal{I}_{ij} := [0, \underline{\delta}_{ij}] \cup [\bar{\delta}_{ij}, 1]$ and $\mathcal{I}_{k\ell} := [0, \underline{\delta}_{k\ell}] \cup [\bar{\delta}_{k\ell}, 1]$. Without loss of generality, let us assume that $\widetilde{\mathcal{G}}_{ij}(\delta_\star) = \widetilde{\mathcal{R}}_i(\delta_\star)$ and $\widetilde{\mathcal{G}}_{k\ell}(\delta_\star) = \widetilde{\mathcal{R}}_k(\delta_\star)$. That is, $\delta_\star = \hat{\delta}_{ik}$ the intersection point of $\widetilde{\mathcal{R}}_i(\delta)$ and $\widetilde{\mathcal{R}}_k(\delta)$ (see Figure 3 in Section 4.3.2). But, from Proposition 5 we have that

$$\widetilde{\mathcal{R}}_i(\delta_\star) < \widetilde{\mathcal{G}}_{ik}(\delta_\star),$$

which again contradicts the optimality of $\widetilde{V}(\delta_\star)$.

From the previous two cases we conclude that the situation in Figure 7 cannot happen at optimality, that is, that $\widetilde{V}(\delta)$ must be smooth in $(0, 1)$. □

Proof of Proposition 8: The result follows from the continuity of the operator defined by the optimization problem (20). For completeness, assume that k is large enough so that $|1 - Q^k(x, \mathcal{E}, \theta)/\mathcal{Q}(x, \mathcal{E})| < \varepsilon$. Observe that

$$\left|\frac{Q^k(x, \mathcal{E}, \theta)}{\mathcal{Q}'(x, \mathcal{E})} - 1\right| \leq \max\left\{\left|\frac{(1-\varepsilon)\,\mathcal{Q}(x, \mathcal{E})}{\mathcal{Q}'(x, \mathcal{E})} - 1\right|, \left|\frac{(1+\varepsilon)\,\mathcal{Q}(x, \mathcal{E})}{\mathcal{Q}'(x, \mathcal{E})} - 1\right|\right\}.$$

Therefore, subject to $\sum_{x\in\mathcal{E}} \mathcal{Q}(x,\mathcal{E}) = 1$, we have that

$$\min_{\mathcal{Q}'\geq 0} \max_{\theta\in\{\theta_0,\theta_1\}} \max_{x\in\mathcal{E}} \left| \frac{Q^k(x,\mathcal{E},\theta)}{\mathcal{Q}'(x,\mathcal{E})} - 1 \right| \leq \min_{\mathcal{Q}'\geq 0} \max_{x\in\mathcal{E}} \max \left\{ \left| \frac{(1-\varepsilon)\,\mathcal{Q}(x,\mathcal{E})}{\mathcal{Q}'(x,\mathcal{E})} - 1 \right|, \left| \frac{(1+\varepsilon)\,\mathcal{Q}(x,\mathcal{E})}{\mathcal{Q}'(x,\mathcal{E})} - 1 \right| \right\} \leq \varepsilon.$$

The second inequality is obtained by taking $\mathcal{Q}' = \mathcal{Q}$. This shows that the optimization operator is continuous at $\mathcal{Q}$ and that $\mathcal{Q}^k \to \mathcal{Q}$ as $k \to \infty$. □

Proof of Proposition 10: Let $\widetilde{\mathcal{E}}^*$ be a solution to (27). We will prove the first part of the proposition by showing that $\widetilde{\mathcal{E}}^*$ satisfies the following properties:

1. If $i \in \widetilde{\mathcal{E}}^*$ and $\Delta u_i \geq \Delta\bar{u}(\widetilde{\mathcal{E}}^*)$ then $j \in \widetilde{\mathcal{E}}^*$ for all $j \geq i$ such that $\Delta u_i < \Delta u_j$.
2. If $i \in \widetilde{\mathcal{E}}^*$ and $\Delta u_i \leq \Delta\bar{u}(\widetilde{\mathcal{E}}^*)$ then $j \in \widetilde{\mathcal{E}}^*$ for all $j \leq i$ such that $\Delta u_j < \Delta u_i$.

These two conditions imply that there exist two integers n_1 and n_2 such that $\widetilde{\mathcal{E}}^* = \mathcal{E}[n_1, n_2]$.
We will only show the first point since the second follows the same line of arguments. Suppose by contradiction that there exist $i \in \widetilde{\mathcal{E}}^*$ and $j \notin \widetilde{\mathcal{E}}^*$ such that $\Delta\bar{u}(\widetilde{\mathcal{E}}^*) \leq \Delta u_i < \Delta u_j$. Let us consider another display set $\widehat{\mathcal{E}} = \widetilde{\mathcal{E}}^* \cup \{j\} \setminus \{i\}$. We will show that $\tilde{\sigma}^2(\widehat{\mathcal{E}}) > \tilde{\sigma}^2(\widetilde{\mathcal{E}}^*)$ which contradicts the optimality of $\widetilde{\mathcal{E}}^*$. Let $m = m_{\widetilde{\mathcal{E}}^*} = m_{\widehat{\mathcal{E}}}$ denote the cardinality of the sets $\widetilde{\mathcal{E}}^*$ and $\widehat{\mathcal{E}}$, we have that

$$\begin{aligned}
\tilde{\sigma}^2(\widetilde{\mathcal{E}}^*) &= \frac{1}{m}\sum_{k\in\widetilde{\mathcal{E}}^*} (\Delta u_k)^2 - \frac{1}{m^2}\Big(\sum_{k\in\widetilde{\mathcal{E}}^*} \Delta u_k\Big)^2 \\
&= \frac{1}{m}\Big(\sum_{k\in\widehat{\mathcal{E}}} (\Delta u_k)^2 + (\Delta u_i)^2 - (\Delta u_j)^2\Big) - \frac{1}{m^2}\Big(\sum_{k\in\widehat{\mathcal{E}}} \Delta u_k + \Delta u_i - \Delta u_j\Big)^2 \\
&= \tilde{\sigma}^2(\widehat{\mathcal{E}}) + \frac{\Delta u_i - \Delta u_j}{m}\left(\Delta u_i + \Delta u_j - \frac{2}{m}\sum_{k\in\widehat{\mathcal{E}}} \Delta u_k - \frac{\Delta u_i - \Delta u_j}{m}\right) \\
&= \tilde{\sigma}^2(\widehat{\mathcal{E}}) + \frac{\Delta u_i - \Delta u_j}{m}\left(\frac{(m-1)\,\Delta u_i + (m+1)\,\Delta u_j}{m} - \frac{2}{m}\Big(\sum_{k\in\widetilde{\mathcal{E}}^*} \Delta u_k + \Delta u_j - \Delta u_i\Big)\right) \\
&= \tilde{\sigma}^2(\widehat{\mathcal{E}}) + 2\,\frac{\Delta u_i - \Delta u_j}{m}\left(\frac{(m+1)\,\Delta u_i + (m-1)\,\Delta u_j}{2\,m} - \frac{1}{m}\sum_{k\in\widetilde{\mathcal{E}}^*} \Delta u_k\right) \\
&= \tilde{\sigma}^2(\widehat{\mathcal{E}}) + 2\,\frac{\Delta u_i - \Delta u_j}{m}\left(\frac{(m+1)\,\Delta u_i + (m-1)\,\Delta u_j}{2\,m} - \Delta\bar{u}(\widetilde{\mathcal{E}}^*)\right) < \tilde{\sigma}^2(\widehat{\mathcal{E}}),
\end{aligned}$$

where the last inequality follows from noticing that the argument inside the large parentheses in the last line is positive since $\Delta\bar{u}(\widetilde{\mathcal{E}}^*) \leq \Delta u_i < \Delta u_j$.
Let us now turn to the proof of second part of the proposition. To this end, let us suppose that all $\{\Delta u_i\}$ are non-negative. (The proof of the case where all $\{\Delta u_i\}$ non-positive uses the same argument.) We will prove the result by invoking the following lemma.

Lemma 4. *Let X a bounded random variable on $[0, A]$. Then $\mathbb{V}\mathrm{ar}[X] \leq A^2/4$.*

Proof of Lemma 4: We prove first the lemma. For this notice that

$$\mathbb{V}\text{ar}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \leq A\,\mathbb{E}[X] - (\mathbb{E}[X])^2 = \mathbb{E}[X](A - \mathbb{E}[X]).$$

The inequality is due to the fact that $X \in [0, A]$. Finally, we observe that $g(x) = x(A - x)$ is maximized on $[0, A]$ at $x = A/2$ with $g(1/2) = A^2/4$. Hence, $\mathbb{V}\text{ar}[X] \leq A^2/4$. □

To use this result, note maximizing the value of $\tilde{\sigma}^2(\mathcal{E})$ over $\mathcal{E}$ is equivalent to maximize the variance of a non-negative random variable $X_{\mathcal{E}}$ taking values in the set $(\Delta u_i \colon i \in \mathcal{E})$ with equal probability. It follows from Lemma 4 that

$$\mathbb{V}\text{ar}[X_{\mathcal{E}}] \leq \frac{1}{4} \max_{i \in \mathcal{E}} \{\Delta u_i^2\} = \frac{(\Delta u_n)^2}{4}.$$

For the second inequality, recall that we have indexed the products so that $\Delta u_1 \leq \Delta u_2 \leq \cdots \leq \Delta u_n$ and we are assuming that they are all non-negative (i.e., $\Delta u_1 \geq 0$). At the same time, if we set $\widetilde{\mathcal{E}}^* = \{0, n\}$, then it is easy to see that

$$\mathbb{V}\text{ar}[X_{\widetilde{\mathcal{E}}^*}] = \frac{(\Delta u_n)^2}{4}.$$

We conclude that $\widetilde{\mathcal{E}}^*$ maximizes $\tilde{\sigma}(\mathcal{E})$ over $\mathcal{E} \in \mathscr{E}$. □

B On the Convexity of the Value Function

In this appendix we discuss how to take advantage of the convexity of the value function (see Proposition 1) to simplify the optimization problem by eliminating some experiments that are dominated by others. This can be particular useful in some applications where the cardinality of the set of possible experiments $\mathscr{E}$ can be rather large adding an extra layer of complexity to the problem of solving the HJB equation in (5). For example, in the context of an optimal assortment selection problem with n products, there are $2^n - 1$ possible display sets that could be offered. To reduce the number of potential experiments to consider we can use the fact that the value function $\Pi(\delta)$ is convex to eliminate those experiments that are dominated in a *convex order dominance* sense.

Indeed, for every $\delta \in (0, 1)$ and $\mathcal{E} \in \mathscr{E}$, let $Z(\delta, \mathcal{E}) := \delta + \eta(\delta, x, \mathcal{E})$ be the random variable that defines the value of the posterior belief when the prior belief is δ and experiment $\mathcal{E}$ is selected. Note that $\mathbb{E}_\delta[Z(\delta, \mathcal{E})] = \delta$ for all $\mathcal{E} \in \mathscr{E}$. Suppose that for two experiments $\mathcal{E}_1, \mathcal{E}_2 \in \mathscr{E}$ we have that $Z(\delta, \mathcal{E}_1) \leq_{cx} Z(\delta, \mathcal{E}_2)$, that is, the random variable $Z(\delta, \mathcal{E}_2)$ dominates $Z(\delta, \mathcal{E}_1)$ in the convex order sense (see Shaked and Shanthikumar, 1994). Then, by convexity of Π in Proposition 1, we get that $\mathbb{E}_\delta[\Pi(Z(\delta, \mathcal{E}_1)] \leq \mathbb{E}_\delta[\Pi(Z(\delta, \mathcal{E}_2)]$. As a result, experiment $\mathcal{E}_1$ can be excluded from the set of possible experiments to be considered when the belief process equals δ. It is worth highlighting that this elimination procedure does not rely on any specific knowledge of the value function beyond the fact that it is convex.

One class of problems for which this elimination scheme is particularly simple is the class of problems in which each experiment can only generate two possible outcomes, that is, $|\mathcal{X}_{\mathcal{E}}| = 2$ for all $\mathcal{E} \in \mathscr{E}$. This is

an important special case given the popularity of pairwise comparison methods (see Szörényi et al., 2015, Heckel et al., 2019 and references therein). In this case, $Z(\delta, \mathcal{E}_1) \leq_{cx} Z(\delta, \mathcal{E}_2)$ if and only if the range of $Z(\delta, \mathcal{E}_1)$ is contained in the range of $Z(\delta, \mathcal{E}_2)$. But this is the same as requiring that the range of the likelihood ratio $\mathcal{L}(\mathcal{E}_1)$ is contained in the range of $\mathcal{L}(\mathcal{E}_2)$, which is a condition that is independent of δ and one can check efficiently. For example, if we apply this elimination scheme to the special instance in Example 1, we get that Experiments 1, 8 and 9 can be eliminated. To see this, note that the likelihood ratio of Experiment 1 takes values $\mathcal{L}(\mathcal{E}_1) \in \{0.3, 1.078\}$ while the likelihood ratio for Experiment 2 takes values $\mathcal{L}(\mathcal{E}_2) \in \{0.2, 1.2\}$. It follows that Experiment 2 dominates Experiment 1. (Similar calculations reveal that Experiments 7 dominates Experiments 8 and 9.)

We can use stochastic dominance one step further to provide some additional intuition for the optimality of our proposed maximum volatility policy. Indeed, since $Z(\delta, \mathcal{E}_2) \leq_{cx} Z(\delta, \mathcal{E}_1)$ implies that $\mathbb{V}\text{ar}[Z(\delta, \mathcal{E}_2)] \leq \mathbb{V}\text{ar}[Z(\delta, \mathcal{E}_1)]$, we can implement a heuristic policy which for each value of δ selects the experiment $\mathcal{E}^{\text{H}}(\delta)$ that maximizes $\mathbb{V}\text{ar}[Z(\delta, \mathcal{E})]$. But $\mathbb{V}\text{ar}[Z(\delta, \mathcal{E})] = \mathbb{E}[\eta^2(\delta, x, \mathcal{E})]$ and so this heuristic policy reduces to

$$\mathcal{E}^{\text{H}}(\delta) = \underset{\mathcal{E} \in \mathscr{E}}{\text{argmax}} \Big\{ \mathbb{E}\Big[\eta^2(\delta, x, \mathcal{E})\Big] \Big\}.$$

But as we have shown in Section 4, this simple experimentation policy is indeed optimal in the asymptotic regime in which the magnitude of the jumps $\eta^2(\delta, x, \mathcal{E})$ converges uniformly to zero, i.e., in a regime in which each experiment becomes less and less informative.

We can also use the convexity of the value function to gain some intuition about this asymptotic result. Indeed, if we assume that the value function is twice-continuously differentiable, then a second order expansion of $\Pi(\delta)$ leads to the following equality:

$$\Pi\big(\delta + \eta(\delta, x, \mathcal{E})\big) - \Pi(\delta) = \dot{\Pi}(\delta)\, \eta(\delta, x, \mathcal{E}) + \frac{1}{2}\, \ddot{\Pi}(\delta)\, (\eta(x, \delta, \mathcal{E}))^2 + O((\eta(\delta, x, \mathcal{E}))^3).$$

The optimal experiment is characterized in equation (6), which we write here as follows,

$$\mathcal{E}^*(\delta) = \underset{\mathcal{E} \in \mathscr{E}}{\text{argmax}} \Big\{ \mathbb{E}_\delta\Big[\Pi\big(\delta + \eta(\delta, x, \mathcal{E})\big) - \Pi(\delta)\Big] \Big\}.$$

By Lemma 2, δ_t is a martingale and so we have that $\mathbb{E}_\delta[\eta(\delta, x, \mathcal{E})] = 0$. Also, by convexity of the value function, $\ddot{\Pi}(\delta) \geq 0$. Combining these two observations together with the assumption that the magnitude of $\eta(\delta, x, \mathcal{E})$ is uniformly small, we conclude that

$$\mathcal{E}^*(\delta) = \underset{\mathcal{E} \in \mathscr{E}}{\text{argmax}}\, \mathbb{E}_\delta\Big[\eta^2(x, \delta, \mathcal{E})\Big].$$

C Additional Numerical Discussions

Running Times: Another dimension of performance is the computational time required to compute a policy and its corresponding value function. Table 3 shows the average running time (in seconds) of

the optimal, Asymptotic and Maximum Volatility policies, as a function of the number of products n available in the menu of prototypes. As we can see, the time required to compute an optimal solution grows exponentially fast with the number of products while the time needed to compute the Asymptotic or Maximum Volatility solution remains low across the range of values of n considered in Table 3.

Average Running Time (in seconds)

	$n=3$	$n=6$	$n=9$	$n=12$	$n=15$
Optimal	0.90	19.00	208.60	22.20×10^2	24.90×10^3
A	0.10	0.19	0.22	0.22	0.27
MV	0.11	0.22	0.24	0.34	0.97

DATA: $\mu = 1$, $r = 0.05$, $G(\delta) = \max\{6 - 30\,\delta, 4 - 5\,\delta, 3\,\delta, -20 + 25\,\delta\}$ and $\Lambda = 2$, $\mathbb{V}\mathrm{ar}[\varepsilon] = \pi^2/(6\,\mu^2)$.

Table 3: Running times of the optimal, Asymptotic and Maximum Volatility policies.

Value of Optimal Stopping: We continue our numerical experiments investigating the option value that the DM has by being able to stop the experimentation process at an arbitrary time. Our interest in measuring the value of optimal stopping is driven by the fact that most practical implementations of crowdvoting are executed with a fix, predetermined, time horizon and so we are interested in measuring the opportunity costs of these implementations.

To this end, let us compare the expected payoffs that the DM collects if she uses the Maximum Volatility experimentation policy $\mathcal{E}^{\mathrm{MV}}(\delta)$ in equation (22) with and without optimal stopping. For the case with optimal stopping, this expected payoff is Π^{MV} as defined above. For the case without optimal stopping, we assume that the DM has a fixed predetermined "budget of experimentation" of T votes that she can collect. In practice, this budget might reflect external constraints on the amount of time or monetary resources available to experiment. Within this budget of experimentation we assume the DM implements the maximum volatility policy $\mathcal{E}^{\mathrm{MV}}(\delta)$. We denote by $\Pi^{\mathrm{MV}}_{\mathrm{T}}$ the corresponding payoff. We define the value of optimal stopping by

$$\text{Value of Optimal Stopping:} \quad \max_{\delta\in[0,1]} \left\{ \frac{\Pi^{\mathrm{MV}} - \Pi^{\mathrm{MV}}_{\mathrm{T}}}{\Pi^{\mathrm{MV}}} \right\}.$$

Figure 8 illustrates the average value of optimal stopping in a concrete instance of the problem with $n = 5$ products for 100 randomly generated instances in which consumers' utilities $u_0(i)$ and $u_1(i)$ are uniformly distributed in [0,1] for $i \in [n]$. Panel (a) depicts the average value of optimal stopping when experimentation is constrained to last exactly T rounds. On the other hand, panel (b) depicts the average value of optimal stopping when the experimentation is constrained to be at most T rounds, that is, in this case the DM is able to stop experimenting before collecting T votes. Also, for comparison purposes, Figure 8 includes the average value of optimal stopping when the full display rule $\mathcal{E}^{\mathrm{F}}$ in (31) is used.

As we can see from the figure, the value of optimal stopping can be quite significant depending on the value of T. This is specially clear on panel (a) in which the value of optimal stopping can be as large as 20% or more if the number of votes T is too small or too large. Intuitively, when T is too small the DM is not able to collect enough information and ends up making wrong decisions. On the other hand,

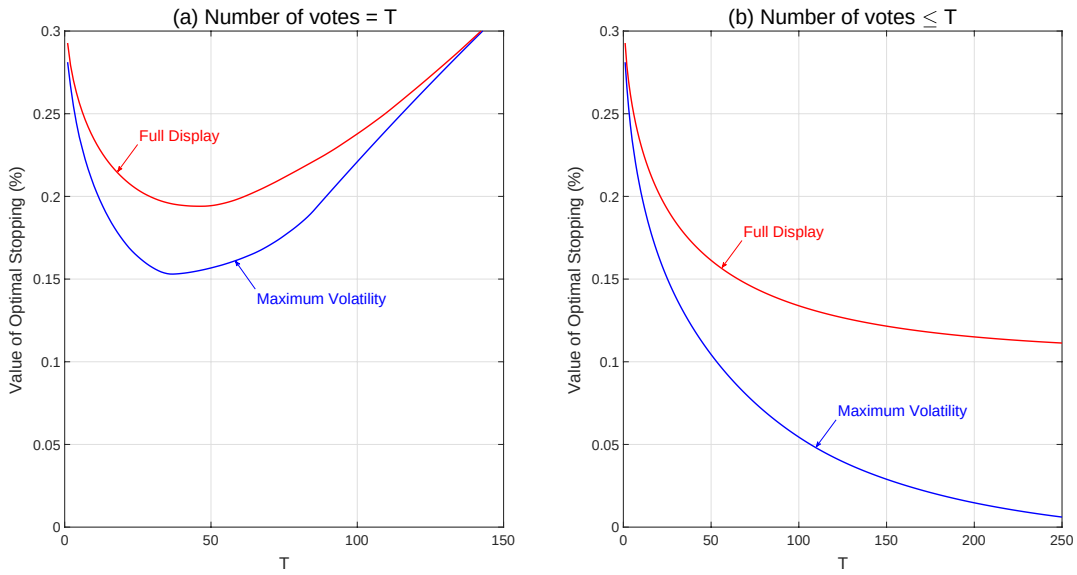

Figure 8: Value of Optimal Stopping as a function of the number of votes T for the maximum volatility and full display strategies. DATA: $\mu = 1$, $r = 0.05$, $G(\delta) = \max\{6 - 30\,\delta, 4 - 5\,\delta, 3\,\delta, -20 + 25\,\delta\}$, $\Lambda = 2$, $\mathbb{V}\text{ar}[\varepsilon] = \pi^2/(6\,\mu^2)$.

when T is too large and the DM exhausts the experimentation budget, then she is guaranteed to collect a large amount of information but pays the price of delaying a final decision too much, which again has a negative effect on payoffs because of discounting, i.e., the DM collects more information that needed. For panel (b), as expected, the value of optimal stopping decreases monotonically with T as the DM in this case is not forced to exhaust all her experimentation budget.

In this example, the average value of optimal stopping under a maximum volatility experimentation rule is minimized around $T = 40$ when the DM operates under the constraint of collecting exactly T votes (panel a) and is about 15%. In contrast, if a Full display policy is used under the same constraint the value of optimal stopping is significantly higher, achieving a minimum around 18% when $T = 45$ votes. By comparing panels (a) and (b) we can appreciate the option value of optimal stopping; not only the value of optimal stopping is monotonically decreasing in T on panel (b) but is also significantly smaller compared to panel (a). These results underscore the significance of giving the DM the option to stop at any time as well as the benefits –from a learning perspective– of using maximum volatility $\mathcal{E}^{\text{MV}}(\delta)$ instead of the popular full display strategy.

Comparison to MNL Bandit Algorithms: We conclude our numerical experiment by comparing our proposed Maximum Volatility policy to a couple of policies from the growing literature on multi-armed bandit problems. We specifically selected one MNL-bandit algorithm and one best arm identification algorithm. The reason for selecting algorithms from this part of the broad literature on sequential testing is that we can cast our assortment selection model in Section 6 as a multi-armed bandit, where each assortment can be viewed as an arm and where at each arrival the DM has to pull one of them to experiment with. Moreover, the growing literature on bandit problems and specifically MNL-bandit setups have considered assortment planning as one of their primary and most natural application (see, e.g. Caro and Gallien (2007) and Agrawal et al. (2019)). Many of the algorithms in this literature have

been developed with the objective of minimizing the DM regret over a finite time horizon. This setting is different than ours in the sense that our objective is to identify, as quick as possible, the best possible "arm" (action) to choose. However, we can still adapt our proposed Maximum Volatility methodology to this minimum regret setting. This shouldn't be of great concern given that our approach is obtained for a general reward function and as discussed our experimentation policy is independent of the duration.

To this end, we assume that the DM has a non-informative uniform prior (i.e., $\delta = 0.5$) and uses the MV policy for a fixed number of votes T. Using a slight abuse of notation, let us denote by δ_T^{MV} the DM's posterior belief after this voting period has ended and let $a_T^{\text{MV}} \in \mathcal{A}^*(\delta_T^{\text{MV}})$ be the optimal action she chooses. For simplicity, we will consider the case in which the optimal action sets $\mathcal{A}^*(\delta)$ are restricted to include a single product, in other words, the DM wants to launch a single product into the marketplace. We let $\mathcal{R}(a_T^{\text{MV}}, \Theta)$ be the reward associated with this policy as a function of Θ. On the other hand, we define $\mathcal{R}^*(\Theta) = \max_{a \in \mathscr{A}}\{\mathcal{R}(a, \Theta)\}$ to be the optimal reward of a clairvoyant who knows the true value of Θ. The terminal regret under this modified MV policy is given by $\Delta\mathcal{R}^{\text{MV}} := \mathcal{R}^*(\Theta) - \mathcal{R}(a_T^{\text{MV}}, \Theta)$.

The following are the two alternative algorithms that we use for comparison:

- MNL-Bandit. The first algorithm that we consider is the one proposed by Agrawal et al. (2019) (Algorithm 1). This is a 'general purpose' algorithm that makes no prior assumption on the MNL model, except for requiring that the no purchase option is the most frequent choice. The algorithm is also designed with the objective of minimizing the rate at which cumulative regret grows as a function of the number of votes T rather than the terminal regret at T, so the comparison is not ideal. The MNL-Bandit is a UCB-type algorithm that periodically during the voting process estimates upper bounds on the attraction scores $v_i = \exp(\mu\, u_i)$ of each product $i \in [n]$ and uses these upper bounds to display the assortment that maximizes rewards. In the implementation of the MNL-Bandit algorithm, we initialize the value of the attraction scores to one. We let v_{iT} denote the terminal estimate of the attraction score for product i after T votes. Using these terminal scores, we define $a_T^{\text{MNL-B}}$ to be the action (product) that maximizes the DM expected reward. The terminal regret of thr MNL-Bandit algorithm is given by $\Delta\mathcal{R}^{\text{MNL-B}} := \mathcal{R}^*(\Theta) - \mathcal{R}(a_T^{\text{MNL-B}}, \Theta)$.

- Top-Two Probability Sampling (TTPS). This algorithm is a variation of a recently proposed algorithm by Russo (2020) for best arm identification. Like MV, TTPS is a Bayesian algorithm that updates the belief δ after each vote. The key difference is in the experiment that is used at every voting epoch. For each value of δ, TTPS identifies the best and second best experiments, in terms of the reward they generate, and selects one of them at random with probabilities β and $1 - \beta$, respectively, where β is a tuning parameter. In our simulations we use $\beta = 0.5$, which is the default value used by Russo (2020). We let δ_T^{TTPS} denote the posterior belief produced by the TTPS algorithm after T votes and let $a_T^{\text{TTPS}} \in \mathcal{A}^*(\delta_T^{\text{TTPS}})$ be the corresponding optimal action. The terminal regret of TTPS is equal to $\Delta\mathcal{R}^{\text{TTPS}} := \mathcal{R}^*(\Theta) - \mathcal{R}(a_T^{\text{TTPS}}, \Theta)$.

In our numerical experiments we use simulation to evaluate the values of $\Delta\mathcal{R}^{\text{MV}}, \Delta\mathcal{R}^{\text{MNL-B}}$ and $\Delta\mathcal{R}^{\text{TTPS}}$. Figure 9 depicts the average terminal regret of these three policies for values of T ranging from 100 to one million votes for a specific instance with $n = 5$ products. The attraction scores $v_i(\theta) = \exp(\mu\, u_i(\theta))$ and per unit margin $p_i - c_i$ for each of the five products is reported in Table 4.

Product	1	2	3	4	5
$v_i(\theta_0)$	0.05	0.08	0.012	0.05	0.04
$v_i(\theta_1)$	0.032	0.07	0.018	0.12	0.043
$p_i - c_i$	210	121.5	506	42	208

Table 4: Vectors of attraction scores and margins for the instance used in the computational experiments reported in Figure 9.

For each algorithm and value of T, we run 1,000 simulations to compute the average terminal regret. Figure 9 also depicts the 95% confidence interval of the mean terminal regret (error bars).

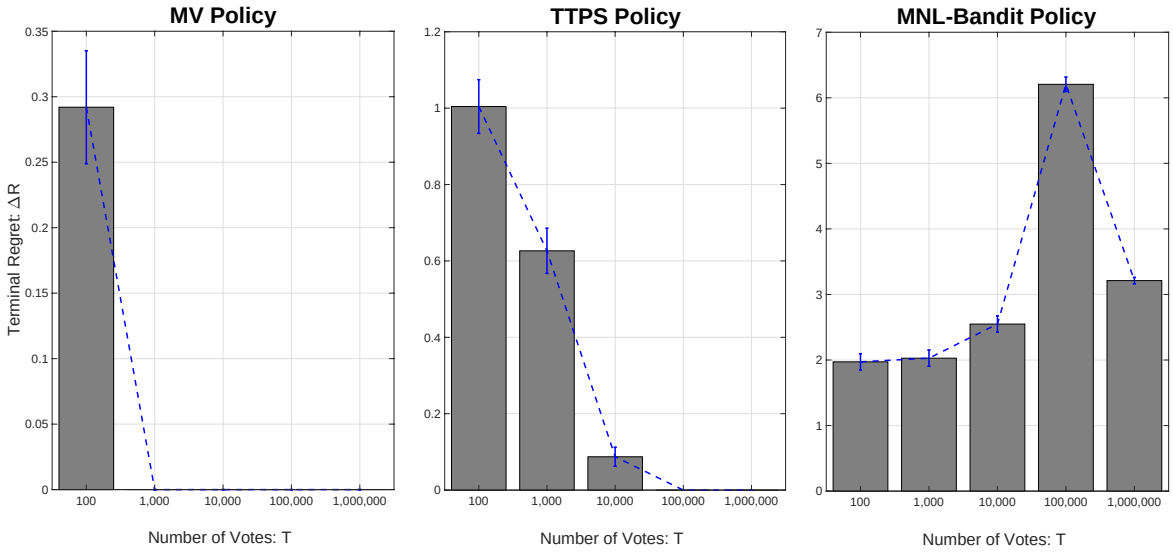

Figure 9: Average terminal regret for the MV, TTPS and MNL-Bandit algorithms as a function of the number of votes T. For each value of T, the average is calculated over 1,000 simulations. The error bars indicate the 95% confidence interval for the mean.

As we can see form the figure, the MV Policy outperforms the other two in terms of achieving a lower terminal regret with significantly fewer votes. Indeed, for $T \geq 1,000$, the MV policy has essentially zero terminal regret. On the other hand, the TTPS policy needs $T \geq 100,000$ to achieve a zero terminal regret. Finally, the MNL-Bandit algorithm does not produce a zero terminal regret for any value of T.

We note that we need to read the results in Figure 9 with cautious. The fact that the MNL-Bandit algorithm does not perform well in terms of minimizing terminal regret should not be surprising as this policy is not designed for this purpose but rather to minimize cumulative regret. To provide a complete picture of the performance of these policies, we have also run a set of experiments to measure their cumulative regret as a function of T.

Figure 10 depicts the average (per vote) cumulative regret of the three policies. As we can see, only the MNL Bandit algorithm achieves a sublinear regret in T while both MV and TTPS have linear cumulative regret. Again, this should be expected since MV and TTPS are pure learning policies designed to identify

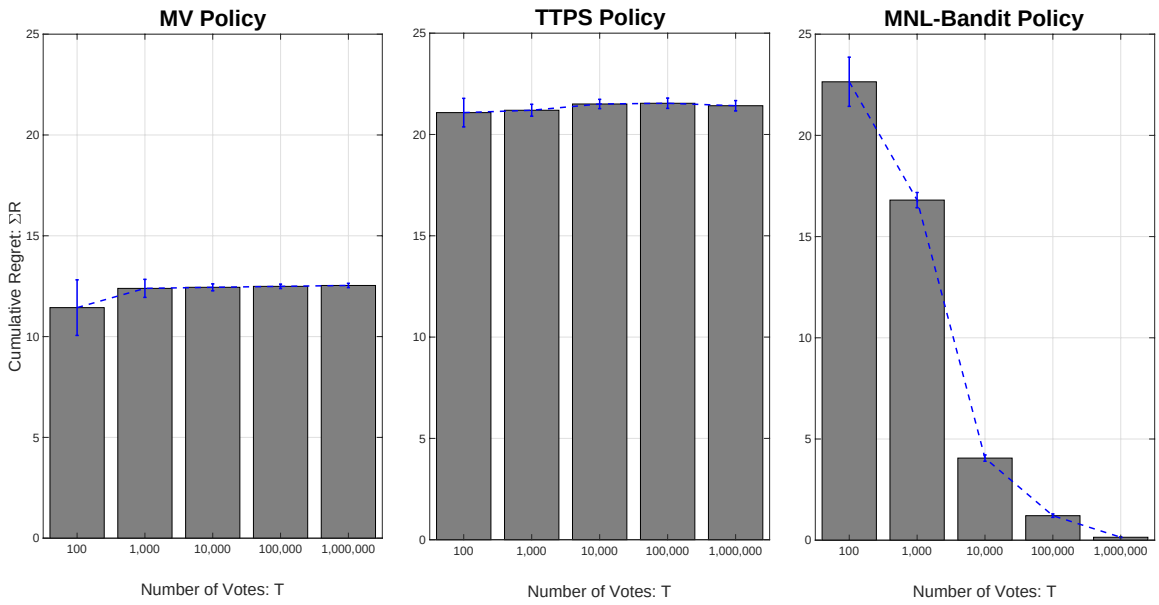

Figure 10: Average cumulative regret for the MV, TTPS and MNL-Bandit algorithms as a function of the number of votes T. For each value of T, the average is calculated over 1,000 simulations. The error bars indicate the 95% confidence interval for the mean.

as quickly as possible the best assortment.