A DEFERRED PROOFS

A.1 PROPERTIES OF LOTKA-VOLTERRA SYSTEMS

Definition A.1 (Static policy equilibrium)**.** *Let $\lambda(t), q(t)$ denote a Lokta-Volterra model characterized by parameters $\theta = (\alpha, \beta, \gamma, \delta) \in \mathbb{R}^4_+$, as defined in Equation 2. Let $p \in [0, 1]$, and denote by π_p the static policy corresponding to p. For $\lambda(0), q(0) > 0$, the static equilibrium of the system is defined as:*

$$\lambda^*(p; \theta) = \lim_{t \to \infty} \lambda(t)$$
$$q^*(p; \theta) = \lim_{t \to \infty} q(t)$$

We denote $\lambda^*(p) = \lambda^*(p; \theta)$ when θ is clear from the context. We denote $\lambda^*(p; u) = \lambda^*(p; \theta_u)$ when a user $u \in \mathcal{U}$ characterized by parameters θ_u is given and clear from the context.

Proposition A.1 (Global stability)**.** *$\lambda^*(p; \theta)$ exists and uniquely defined for all $\theta \in \mathbb{R}^4_+$, $p \in [0, 1]$ and for all initial conditions $\lambda(0), q(0) > 0$.*

Proof. See (Takeuchi, 1996, Section 3.2). □

Lemma A.1 (Equilibrium of LV behavioral model. Formal proof of Lemma 1)**.** *Assume a Lokta-Volterra model characterized by $\theta = (\alpha, \beta, \gamma, \delta) \in \mathbb{R}^4_+$, and let $p \in [0, 1]$ denote the proportion of interactions in which a forced break is served. The static equilibrium of the model under static policy π_p is given by:*

$$\lambda^*(p) = \begin{cases} \frac{\gamma}{\delta} \frac{1}{1-p} \left(1 - \frac{\alpha}{\beta} \frac{1}{1-p}\right) & p \in \left[0, 1 - \frac{\alpha}{\beta}\right] \\ 0 & \text{otherwise} \end{cases}$$
$$q^*(p) = \begin{cases} \frac{\alpha}{\beta} \frac{1}{1-p} & p \in \left[0, 1 - \frac{\alpha}{\beta}\right] \\ 1 & \text{otherwise} \end{cases}$$

Proof. The LV dynamical system is given by Equation 3:

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = -\alpha\lambda + \beta(1-p)\lambda q$$
$$\frac{\mathrm{d}q}{\mathrm{d}t} = \gamma q(1-q) - \delta(1-p)\lambda q$$

when $p \in \left[0, 1 - \frac{\alpha}{\beta}\right]$ we equate $\frac{\mathrm{d}\lambda}{\mathrm{d}t} = 0$, $\frac{\mathrm{d}q}{\mathrm{d}t} = 0$ and obtain the result. The solution is guaranteed to be valid, as both $\lambda^*(p) > 0$ and $q^*(p) \in [0, 1]$.

Conversely, when $p \notin \left[0, 1 - \frac{\alpha}{\beta}\right]$, there exists $\epsilon > 0$ such that $\frac{\mathrm{d}}{\mathrm{d}t} \log \lambda < -\epsilon < 0$ for all $\lambda > 0$, $q \in [0, 1]$. From this we obtain that $\log \lambda(t)$ tends towards $-\infty$, and therefore $\lambda(t)$ tends towards 0, and $\lambda^*(p) = 0$ as required. When $\lambda(t)$ is close to zero, the interaction terms vanish in the $\frac{\mathrm{d}q}{\mathrm{d}t}$ equation, and $q(t)$ grows logistically towards 1. □

Proposition A.2 (Equilibrium bounds)**.** *For a Lotka-Volterra model, the static equilibrium $\lambda^*(p)$ is bounded by:*

$$0 \leq \lambda^*(p) \leq \frac{\beta\gamma}{4\alpha\delta}$$

Proof. Denote $x = \frac{1}{1-p}$. From Lemma A.1, for $x \in \left[1, \frac{\beta}{\alpha}\right]$ the equilibrium consumption $\lambda^*(x)$ is given by:

$$\lambda^*(x) = \frac{\gamma}{\delta} x \left(1 - \frac{\alpha}{\beta} x\right)$$

and is zero otherwise. The equilibrium is a quadratic function of x with roots $x \in \left\{0, \frac{\beta}{\alpha}\right\}$, and therefore attains its maximum at $x = \frac{\beta}{2\alpha}$. Plugging back the maximizing x into λ^* we obtain

the upper bound. Lower bound is attained as the equilibrium in Lemma A.1 is clipped by 0 from below. □

Lemma A.2 (Optimal static policy. Formal proof of Lemma 2)**.** *The optimal static policy for a Lotka-Volterra system is given by:*

$$p_{\text{opt}} = \begin{cases} 1 - 2\frac{\alpha}{\beta} & \frac{\alpha}{\beta} \leq \frac{1}{2} \\ 0 & \frac{\alpha}{\beta} > \frac{1}{2} \end{cases}$$

And the optimal equilibrium engagement rate is given by:

$$\lambda^*_{\text{opt}} = \begin{cases} \frac{\beta\gamma}{4\alpha\delta} & \frac{\alpha}{\beta} \leq \frac{1}{2} \\ \frac{\gamma}{\delta}\left(1 - \frac{\alpha}{\beta}\right) & \frac{\alpha}{\beta} > \frac{1}{2} \end{cases}$$

Proof. Denote $x = \frac{1}{1-p}$. From Proposition A.2, the global maximum of $\lambda^*(x)$ is attained at $x = \frac{\beta}{2\alpha}$. Consider two cases: When $\frac{\alpha}{\beta} \leq \frac{1}{2}$, we obtain that $x_{\text{opt}} = \frac{\beta}{2\alpha} \geq 1$, and therefore $p_{\text{opt}} = 1 - \frac{1}{x} \in [0, 1]$. From this we obtain that in this case the global maximum is attained on the simplex, and given by the formula from Proposition A.2. Conversely, when $\frac{\alpha}{\beta} > \frac{1}{2}$, we obtain $p = 1 - \frac{1}{x} < 0$, and therefore x_{opt} translates to a negative value of p. As $\lambda^*(p)$ is uni-modal, the optimal policy restricted to the simplex $[0, 1]$ in this case is attained on the closest boundary point $p = 0$.

Figure 2 provides graphical intuition for this proof (left and center subplots). □

Proposition A.3 (Inference of α/β from two-treatment equilibrium data. Formal proof of Proposition 1)**.** *Let $\lambda(t), q(t)$ be a Lokta-Volterra model, let $p_1, p_2 \in [0, 1]$. Denote by $\lambda^*(p_1), \lambda^*(p_2)$ the static equilibrium rates corresponding to static policies π_{p_1}, π_{p_2}, and assume $\lambda^*(p_1), \lambda^*(p_2) > 0$. The parameter ratio $\frac{\alpha}{\beta}$ is given by the following formula:*

$$\frac{\alpha}{\beta} = \frac{(1-p_2)\lambda^*(p_2) - (1-p_1)\lambda^*(p_1)}{\frac{1}{1-p_1} - \frac{1}{1-p_2}}$$

Proof. From Lemma A.1, the equilibrium consumption $\lambda^*(p)$ is given by:

$$\begin{aligned} \lambda^*(p) &= \frac{\gamma}{\delta}\frac{1}{1-p}\left(1 - \frac{\alpha}{\beta}\frac{1}{1-p}\right) \\ &= \frac{\gamma}{\delta}\frac{1}{1-p} - \frac{\alpha}{\beta}\frac{\gamma}{\delta}\left(\frac{1}{1-p}\right)^2 \end{aligned}$$

When $\lambda^*(p_i)$ is observed for different policies $p_1, \ldots, p_m \in \left[0, 1 - \frac{\alpha}{\beta}\right]$, we obtain a polynomial regression problem for the parameters $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta}\frac{\gamma}{\delta}$, which can be solved e.g using Non-Negative Least Squares.

When $m = 2$, we obtain a system of two linear equations. Apply Cramer's rule to obtain:

$$\frac{\gamma}{\delta} = \frac{\frac{\lambda^*(p_2)}{(1-p_1)^2} - \frac{\lambda^*(p_1)}{(1-p_2)^2}}{\frac{1}{(1-p_1)(1-p_2)^2} - \frac{1}{(1-p_1)^2(1-p_2)}} = \frac{(1-p_2)^2\lambda^*(p_2) - (1-p_1)^2\lambda^*(p_1)}{p_2 - p_1} \tag{11}$$

$$\frac{\alpha}{\beta}\frac{\gamma}{\delta} = \frac{\frac{\lambda^*(p_2)}{(1-p_1)} - \frac{\lambda^*(p_1)}{(1-p_2)}}{\frac{1}{(1-p_1)(1-p_2)^2} - \frac{1}{(1-p_1)^2(1-p_2)}} = (1-p_1)(1-p_2)\frac{(1-p_2)\lambda^*(p_2) - (1-p_1)\lambda^*(p_1)}{p_2 - p_1} \tag{12}$$

And therefore $\frac{\alpha}{\beta}$ is given by:

$$\frac{\alpha}{\beta} = \frac{\frac{\lambda^*(p_2)}{(1-p_1)} - \frac{\lambda^*(p_1)}{(1-p_2)}}{\frac{\lambda^*(p_2)}{(1-p_1)^2} - \frac{\lambda^*(p_1)}{(1-p_2)^2}} = (1-p_1)(1-p_2)\frac{(1-p_2)\lambda^*(p_2) - (1-p_1)\lambda^*(p_1)}{(1-p_2)^2\lambda^*(p_2) - (1-p_1)^2\lambda^*(p_1)}$$

□

A.2 MODEL FITTING FROM ENGAGEMENT PREDICTIONS

Notations. In this section only, we use the common notation $q = 1 - p$ to denote complementary probabilities.

Definition A.2 (Empirical value of α/β)**.** *For single-channel experiments with forced-break probabilities p_1, p_2, denote $\lambda_i = \lambda^*(p_i)$, $f_i = f_{p_i}(u)$, $q_i = 1-p_i$. The empirical value of the $\frac{\alpha}{\beta}$ parameter is given by the following formula:*

$$\frac{\hat{\alpha}}{\beta} = \frac{q_1 q_2 \left(q_1 f_1 - q_2 f_2\right)}{q_1^2 f_1 - q_2^2 f_2}$$

Proposition A.4 (α/β estimation error from prediction errors)**.** *Given a single-channel Lokta-Volterra system with parameter $\frac{\alpha}{\beta} \geq 1$. Let $p_1, p_2 \in \left[1, \frac{\alpha}{\beta}\right]$, denote $\lambda_i^* = \lambda^*(p_i) \in \mathbb{R}_+$, and let $f_i = \lambda_i^* + \varepsilon_i$ be the predicted engagement rates corresponding to p_1, p_2. When $|\varepsilon_1|, |\varepsilon_2| \leq \varepsilon \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}$, the estimation error is bounded by:*

$$\left|\frac{\alpha}{\beta} - \frac{\hat{\alpha}}{\beta}\right| \leq \frac{\varepsilon}{|p_1 - p_2|}\frac{\beta\delta}{\alpha\gamma}$$

Proof. denote $q_i = 1 - p_i$. The value of $\frac{\alpha}{\beta}$ is given by Proposition A.3:

$$\frac{\alpha}{\beta} = \frac{q_1 q_2 \left(q_1 \lambda_1^* - q_2 \lambda_2^*\right)}{q_1^2 \lambda_1^* - q_2^2 \lambda_2^*}$$

And the estimator for $\frac{\alpha}{\beta}$ is obtained by replacing the true value with their predictions:

$$\begin{aligned}\frac{\hat{\alpha}}{\beta} &= \frac{q_1 q_2 \left(q_1 f_1 - q_2 f_2\right)}{q_1^2 f_1 - q_2^2 f_2} \\ &= \frac{q_1 q_2 \left(q_1(\lambda_1^* + \varepsilon_1) - q_2(\lambda_2^* + \varepsilon_2)\right)}{q_1^2(\lambda_1^* + \varepsilon_1) - q_2^2(\lambda_2^* + \varepsilon_2)}\end{aligned}$$

The estimation error is given by:

$$\begin{aligned}\left|\frac{\alpha}{\beta} - \frac{\hat{\alpha}}{\beta}\right| &= \left|\frac{q_1^2 q_2^2 (q_1 - q_2)(\varepsilon_2 \lambda_1^* - \varepsilon_1 \lambda_2^*)}{(q_1^2 \lambda_1^* - q_2^2 \lambda_2^*)(q_1^2 \lambda_1^* - q_2^2 \lambda_2^* - (q_1^2 \varepsilon_1 - q_2^2 \varepsilon_2))}\right| \\ &= \underbrace{(q_1 q_2)^2}_{\equiv \text{(i)}} \underbrace{\left|\frac{q_1 - q_2}{q_1^2 \lambda_1^* - q_2^2 \lambda_2^*}\right|}_{\equiv \text{(ii)}} \underbrace{\left|\varepsilon_2 \lambda_1^* - \varepsilon_1 \lambda_2^*\right|}_{\equiv \text{(iii)}} \underbrace{\left|\frac{1}{q_1^2 \lambda_1^* - q_2^2 \lambda_2^* - (q_1^2 \varepsilon_1 - q_2^2 \varepsilon_2)}\right|}_{\equiv \text{(iv)}}\end{aligned}$$

We now proceed to bound each factor:

- For (i), the term $(q_1 q_2)^2$ is bounded by 1 since $q_1, q_2 \in [0, 1]$.
- For (ii), the term $\left|\frac{q_1 - q_2}{q_1^2 \lambda_1^* - q_2^2 \lambda_2^*}\right|$ is equal to $\left(\frac{\gamma}{\delta}\right)^{-1}$ by Eq. (11).
- For (iii), from Proposition A.2 we obtain the bound $0 \leq \lambda_i^* \leq \frac{\beta\gamma}{4\alpha\delta}$, and therefore the term $|\varepsilon_2 \lambda_1^* - \varepsilon_1 \lambda_2^*|$ is bounded by $2\left(\frac{\beta\gamma}{4\alpha\delta}\right)\varepsilon = \frac{\beta\gamma}{2\alpha\delta}\varepsilon$.

- For (iv), the term $\left|\frac{1}{q_1^2\lambda_1^*-q_2^2\lambda_2^*-(q_1^2\varepsilon_1-q_2^2\varepsilon_2)}\right|$ is equal to:

$$\begin{aligned}\text{(iv)} &\equiv \left|\frac{1}{q_1^2\lambda_1^*-q_2^2\lambda_2^*-(q_1^2\varepsilon_1-q_2^2\varepsilon_2)}\right| \\ &= \frac{1}{|p_1-p_2|}\left|\frac{q_1^2\lambda_1^*-q_2^2\lambda_2^*-(q_1^2\varepsilon_1-q_2^2\varepsilon_2)}{p_1-p_2}\right|^{-1} \\ &= \frac{1}{|p_1-p_2|}\left|\underbrace{\frac{q_1^2\lambda_1^*-q_2^2\lambda_2^*}{p_1-p_2}}_{\text{Eq. (11)}}-\frac{q_1^2\varepsilon_1-q_2^2\varepsilon_2}{p_1-p_2}\right|^{-1} \\ &= \frac{1}{|p_1-p_2|}\left|\frac{\gamma}{\delta}-\frac{q_1^2\varepsilon_1-q_2^2\varepsilon_2}{p_1-p_2}\right|^{-1}\end{aligned}$$

  Note that $\left|\frac{q_1^2\varepsilon_1-q_2^2\varepsilon_2}{p_1-p_2}\right| \leq \frac{2\varepsilon}{|p_1-p_2|}$. When ε is small enough, and specifically when the bound $\varepsilon \leq \frac{\gamma}{\delta}\frac{|p_1-p_2|}{4}$ holds, we obtain:

$$\left|\frac{\gamma}{\delta}-\frac{q_1^2\varepsilon_1-q_2^2\varepsilon_2}{p_1-p_2}\right|^{-1} \leq \frac{\delta}{\gamma}\left|1-\frac{1}{2}\right|^{-1} \leq 2\frac{\delta}{\gamma}$$

  and therefore:

$$\text{(iv)} \leq \frac{2}{|p_1-p_2|}\frac{\delta}{\gamma}$$

Aggregating results (i)-(iv) above, we obtain the overall bound:

$$\begin{aligned}\left|\frac{\alpha}{\beta}-\frac{\hat{\alpha}}{\beta}\right| &= \underbrace{(q_1q_2)^2}_{\leq 1}\underbrace{\left|\frac{q_1-q_2}{q_1^2\lambda_1^*-q_2^2\lambda_2^*}\right|}_{=\frac{\delta}{\gamma}}\underbrace{\left|\varepsilon_2\lambda_1^*-\varepsilon_1\lambda_2^*\right|}_{\leq\frac{\beta\gamma}{2\alpha\delta}\varepsilon}\underbrace{\left|\frac{1}{q_1^2\lambda_1^*-q_2^2\lambda_2^*-(q_1^2\varepsilon_1-q_2^2\varepsilon_2)}\right|}_{\leq\frac{2}{|p_1-p_2|}\frac{\delta}{\gamma}} \\ &\leq \frac{\varepsilon}{|p_1-p_2|}\frac{\beta\delta}{\alpha\gamma}\end{aligned}$$

□

Proposition A.5 (Cost of α/β estimation error)**.** *Let $\frac{\alpha}{\beta}$ be the engagement ratio parameter of a one-channel Lotka-Volterra system, and let $\left(\hat{\frac{\alpha}{\beta}}\right)$ be an estimate of these parameters. Let λ^*_{opt} be the engagement rate of the optimal static policy, and denote $\lambda^*(x) = \lambda^*\left(\hat{p}(x)\right)$. When $\left|\frac{\alpha}{\beta}-\left(\hat{\frac{\alpha}{\beta}}\right)\right| \leq \min\left\{\frac{\alpha}{2\beta}, 1\right\}$ The price of estimation error is bounded by:*

$$\lambda^*_{\text{opt}}-\lambda^*\left(\left(\hat{\frac{\alpha}{\beta}}\right)\right) \leq \left(\frac{\gamma}{\delta}\right)\min\left\{\left(2\frac{\alpha}{\beta}\right)^{-2}\left|\frac{\alpha}{\beta}-\left(\hat{\frac{\alpha}{\beta}}\right)\right|, \left(4\frac{\alpha}{\beta}\right)^{-1}\right\}$$

Proof. Denote $r = \frac{\alpha}{\beta}, x = \left(\hat{\frac{\alpha}{\beta}}\right)$, and assume without loss of generality that $\frac{\gamma}{\delta} = 1$ and $r \leq 1$. The optimal equilibrium engagement rate is given by:

$$\lambda^*_{\text{opt}} = \begin{cases}\frac{1}{4r} & r \in \left(0, \frac{1}{2}\right] \\ 1-r & r \in \left(\frac{1}{2}, 1\right]\end{cases}$$

The chosen policy $\hat{p}(x)$ is given by:

$$\hat{p}(x) = \begin{cases}1-2x & x \in \left[0, \frac{1}{2}\right] \\ 0 & \text{otherwise}\end{cases}$$

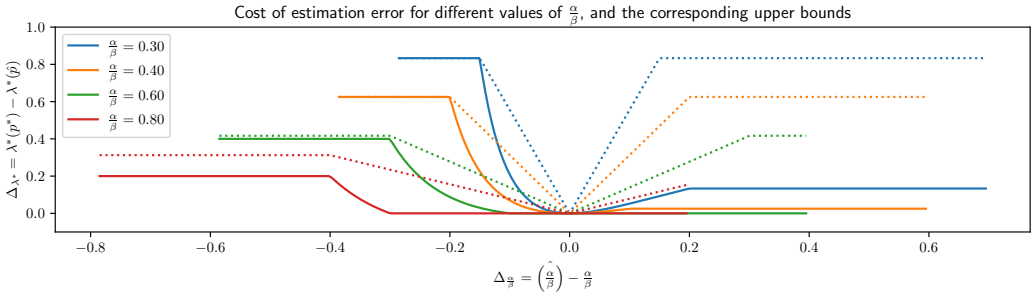

Figure 4: Graphical illustration of Proposition A.5. Cost of estimation error for different values of $\frac{\alpha}{\beta}$, and their corresponding upper bounds given by the claim.

Assume without loss of generality that $x \in \left[0, \frac{1}{2}\right]$, as values of x outside the interval can be clipped to its edges without affecting the result. The equilibrium engagement rate of the selected policy is given by:

$$\lambda^*(x) = \lambda^*\left(\hat{p}(x)\right) = \begin{cases} 0 & x \in \left[0, \frac{r}{2}\right] \\ \frac{1}{2x}\left(1 - \frac{r}{2x}\right) & x \in \left(\frac{r}{2}, \frac{1}{2}\right] \end{cases}$$

Denote $\Delta(x) = \lambda^*_{\text{opt}} - \lambda^*(x)$. We obtain:

$$\Delta(x) = \lambda^*_{\text{opt}} - \lambda^*(x) = \begin{cases} \frac{1}{4r} & r \in \left(0, \frac{1}{2}\right], x \in \left[0, \frac{r}{2}\right] \\ \frac{(x-r)^2}{4x^2 r} & r \in \left(0, \frac{1}{2}\right], x \in \left(\frac{r}{2}, \frac{1}{2}\right] \\ (1-r) & r \in \left(\frac{1}{2}, 1\right], x \in \left[0, \frac{r}{2}\right] \\ (1-r) - \frac{1}{2x}\left(1 - \frac{r}{2x}\right) & r \in \left(\frac{1}{2}, 1\right], x \in \left(\frac{r}{2}, \frac{1}{2}\right] \end{cases}$$

Observe that $\frac{1}{4r} \geq 1 - r$ for all $r \in (0, 1]$, and therefore we obtain for all x, r:

$$\Delta(x) \leq \frac{1}{4r} \tag{13}$$

From the convexity of $\Delta(x)$ in the region around $x = r$ we obtain:

$$\Delta(x) \leq \frac{1}{2r^2}|x - r| \tag{14}$$

Finally, combining the two bounds yields the final result. A geometric interpretation of this claim is illustrated in Figure 4. □

A.3 Optimal stationary policy from engagement predictions

Definition A.3 (Expected observable rate)**.** *Let $u \in \mathcal{U}$, $p \in [0, 1]$, and $T > 0$. Let $p \in [0, 1]$, denote the corresponding static policy by π_p. The expected observable rate $\bar{\lambda}_u(p; T)$ is defined as:*

$$\bar{\lambda}_u(p; u) = \mathbb{E}_\pi\left[\frac{1}{T}\left|\text{TPP}_{\pi_p}(u; T)\right|\right]$$

where expectation is taken over the stochastic decisions of π_p.

Definition A.4 (Lokta-Volterra approximation of TPP)**.** *Let $u \in \mathcal{U}$, and $T > 0$. Denote by p^* the maximizer of expected observable rate:*

$$p^* = \arg\max_{p \in [0,1]} \bar{\lambda}_u(p; u)$$

The LV approximation of $\text{TPP}(u; T)$ is defined as:

$$\theta^*_u = \arg\min_\theta \max_{p \in [0,1]} \left|\bar{\lambda}_u(p; u) - \lambda^*(p; \theta)\right|$$

such that $\arg\max_p \lambda^*(p; \theta) = p^*$. The corresponding approximation error is defined as:

$$\varepsilon_{\text{LV},u} = \max_{p \in [0,1]} \left|\bar{\lambda}_u(p; u) - \lambda^*(p; \theta^*_u)\right|$$

Notations. When u is clear from the context, we denote $\theta^* = \theta^*_u$, $\varepsilon_{\mathrm{LV}} = \varepsilon_{\mathrm{LV},u}$. We use $\alpha^*, \beta^*, \ldots$ to refer to the corresponding parts of the Lokta-Volterra parameters vector θ^*.

We are now ready to state and prove the main theorem for this section:

Theorem A.1 (Regret bound for learned static policy. Formal version of Theorem 1)**.** *Let $p_1, p_2 \in [0,1]$ denote two static forced-break policies, and denote by $\mathcal{U}$ the set of users, and assume they remain engaged under the stationary policies $\pi(p_1)$ and $\pi(p_2)$. Assume $S_u(p;T) \sim \mathrm{TPP}_{\pi_p \circ \psi}(u;T)$, and let $\mu = \left(\max_{u\in\mathcal{U}} \frac{\bar{\gamma}_u}{\bar{\delta}_u}\right) \cdot \left(\max_{u'\in\mathcal{U}} \frac{\bar{\delta}_{u'}}{\bar{\gamma}_{u'}}\right)$, $\nu = \max_{u\in\mathcal{U}} \left(\frac{\bar{\beta}_u}{\bar{\alpha}_u}\right)$.*

Let $f_{p_1}, f_{p_2} : \mathcal{U} \to \mathbb{R}_+$ be functions predicting $\frac{1}{T}|S_u(p_1;T)|$, $\frac{1}{T}|S_u(p_2;T)|$, respectively. Denote the learned policy by $\hat{p}$, and the optimal policy by p^*.

If (i) the expected RMSE of f_{p_1}, f_{p_2} is bounded by $\varepsilon_{\mathrm{pred}}$, (ii) the average absolute deviation of $\frac{1}{T}|\mathrm{TPP}(u;T)|$ is bounded by $\varepsilon_{\mathrm{dev}}$, and (iii) the expected LV approximation error of the system is bounded by $\varepsilon_{\mathrm{LV}}$, then the learned policy $\hat{p}$ has bounded regret:

$$\mathbb{E}_{u,\pi}\left[\left|\tfrac{1}{T}|S_u(p^*;T)| - \tfrac{1}{T}|S_u(\hat{p};T)|\right|\right] \le \frac{\eta_{\mathrm{TPP}}}{|p_1 - p_2|}(\varepsilon_{\mathrm{pred}} + \varepsilon_{\mathrm{dev}} + \varepsilon_{\mathrm{LV}})$$

where expectation is taken over stochastic choices of policies, and $\eta_{\mathrm{TPP}} = g(\mu,\nu) \in \mathrm{poly}(\mu,\nu)$.

Proof. By assumption (i), the functions f_{p_1}, f_{p_2} have bounded expected RMSE:

$$\mathbb{E}_u\left[\left(f_{p_i}(u) - \tfrac{1}{T}|S_u(p_i;T)|\right)^2\right] \le \varepsilon^2_{\mathrm{pred}} \tag{15}$$

Applying Jensen's inequality with the convex function $\varphi(x) = x^2$ yields:

$$\left(\mathbb{E}_u\left[\left|f_{p_i}(u) - \tfrac{1}{T}|S_u(p_i;T)|\right|\right]\right)^2 \le \mathbb{E}_u\left[\left(f_{p_i}(u) - \tfrac{1}{T}|S_u(p_i;T)|\right)^2\right]$$

Combining with Eq. (15) and taking the square root, we obtain an upper bound on the expected absolute error:

$$\mathbb{E}_u\left[\left|f_{p_i}(u) - \tfrac{1}{T}|S_u(p_i;T)|\right|\right] \le \varepsilon_{\mathrm{pred}} \tag{16}$$

Let $\Delta_f = |f_{p_i}(u) - \lambda^*(p_i)|$ apply the triangle inequality to obtain:

$$\begin{aligned}\Delta_f &= |f_{p_i}(u) - \lambda^*(p_i)| \\ &\le \left|f_{p_i}(u) - \tfrac{1}{T}|S_u(u;T)|\right| + \left|\tfrac{1}{T}|S_u(u;T)| - \bar{\lambda}(p_i;u)\right| + \left|\bar{\lambda}(p_i;u) - \lambda^*(p_i)\right|\end{aligned}$$

Denote $\varepsilon_f = \varepsilon_{\mathrm{pred}} + \varepsilon_{\mathrm{dev}} + \varepsilon_{\mathrm{LV}}$. Applying the triangle inequality and using the bounds in Eq. (16) together with assumptions (ii), (iii), we obtain:

$$\begin{aligned}\mathbb{E}_{u,\pi}[\Delta_f] \le& \mathbb{E}_u\left[\left|f_{p_i}(u) - \tfrac{1}{T}|S_u(u;T)|\right|\right] \\ &+ \mathbb{E}_{u,\pi}\left[\left|\tfrac{1}{T}|S_u(u;T)| - \bar{\lambda}(p_i;u)\right|\right] \\ &+ \mathbb{E}_u\left[\left|\bar{\lambda}(p_i;u) - \lambda^*(p_i;\theta^*_u)\right|\right] \\ \le& \varepsilon_{\mathrm{pred}} + \varepsilon_{\mathrm{dev}} + \varepsilon_{\mathrm{LV}} = \varepsilon_f\end{aligned} \tag{17}$$

Denote $\theta^*_u = (\alpha,\beta,\gamma,\delta)$. The empirical value $\left(\frac{\hat{\alpha}}{\hat{\beta}}\right)$ of $\left(\frac{\alpha}{\beta}\right)$ is given by Definition A.2. Denote the estimation error by $\Delta_{\frac{\alpha}{\beta}} = \left|\left(\frac{\hat{\alpha}}{\hat{\beta}}\right) - \left(\frac{\alpha}{\beta}\right)\right|$.

By Proposition A.4, the following pointwise upper bound on $\Delta_{\frac{\alpha}{\beta}}$ applies when $\Delta_f \le \frac{\gamma}{\delta}\frac{|p_1-p_2|}{4}$:

$$\Delta_{\frac{\alpha}{\beta}} \le \frac{\Delta_f}{|p_1 - p_2|}\frac{\beta\delta}{\alpha\gamma} \tag{18}$$

Plugging in the bound on the expected value of Δ_f into Eq. (18), we obtain in expectation:

$$\begin{aligned}\mathbb{E}_{u,\pi}\left[\Delta_{\frac{\alpha}{\beta}} \mid \Delta_f \le \frac{\gamma}{\delta}\frac{|p_1-p_2|}{4}\right] &\le \mathbb{E}_{u,\pi}\left[\frac{\Delta_f}{|p_1-p_2|}\frac{\beta\delta}{\alpha\gamma} \mid \Delta_f \le \frac{\gamma}{\delta}\frac{|p_1-p_2|}{4}\right] \\ &\le \frac{\varepsilon_f}{|p_1-p_2|}\max_u \frac{\beta\delta}{\alpha\gamma}\end{aligned} \tag{19}$$

Next, we apply Proposition A.5. Denote $\Delta_{\lambda^*} = \lambda^*(p^*) - \lambda^*(\hat{p})$, and define the following probability event:

$$A = \left(\Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right) \text{ and } \left(\Delta_{\frac{\alpha}{\beta}} \leq \frac{1}{2\nu}\right)$$

Note that the bound in Proposition A.5 is represented as a minimum between two functions, one linear in ε and one constant. To leverage this property, apply the law of total expectation:

$$\mathbb{E}_{u,\pi}[\Delta_{\lambda^*}] = \mathbb{E}_{u,\pi}[\Delta_{\lambda^*} \mid A]\mathbb{P}[A] + \mathbb{E}_{u,\pi}\left[\Delta_{\lambda^*} \mid \bar{A}\right]\mathbb{P}\left[\bar{A}\right] \tag{20}$$

Under A, the first term in Eq. (20) can be bounded by the linear term in Proposition A.5. Taking $\mathbb{P}[A] \leq 1$ and combining with equation Eq. (18):

$$\begin{aligned}
\mathbb{E}_{u,\pi}[\Delta_{\lambda^*} \mid A]\mathbb{P}[A] &\leq \mathbb{E}_{u,\pi}[\Delta_{\lambda^*} \mid A] \\
&\leq \mathbb{E}_{u,\pi}\left[\frac{\beta^2\gamma}{2\alpha^2\delta}\Delta_{\frac{\alpha}{\beta}} \mid A\right] \\
&\leq \mathbb{E}_{u,\pi}\left[\frac{\beta^2\gamma}{2\alpha^2\delta}\frac{\Delta_f}{|p_1 - p_2|}\frac{\beta\delta}{\alpha\gamma} \mid A\right] \\
&\leq \frac{\nu^3}{2|p_1 - p_2|}\varepsilon_f
\end{aligned} \tag{21}$$

The expectation factor in the second term of Eq. (20) can be bounded by the constant term in Proposition A.5:

$$\mathbb{E}_{u,\pi}\left[\Delta_{\lambda^*} \mid \bar{A}\right] \leq \frac{1}{4}\max_u \frac{\beta\gamma}{\alpha\delta} \leq \frac{\nu}{4}\max_u \frac{\gamma}{\delta} \tag{22}$$

Decompose the probability factor $\mathbb{P}\left[\bar{A}\right]$ using the law of total probability:

$$\begin{aligned}
\mathbb{P}\left[\bar{A}\right] &= \mathbb{P}\left[\Delta_f > \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] + \mathbb{P}\left[\Delta_{\frac{\alpha}{\beta}} > \frac{1}{2\nu} \mid \Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right]\mathbb{P}\left[\Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] \\
&\leq \mathbb{P}\left[\Delta_f > \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] + \mathbb{P}\left[\Delta_{\frac{\alpha}{\beta}} > \frac{1}{2\nu} \mid \Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right]
\end{aligned}$$

Apply Markov's inequality $\mathbb{P}[|X| \geq a] \leq \frac{\mathbb{E}[|X|]}{a}$ on the probabilities to obtain:

$$\begin{aligned}
\mathbb{P}\left[\Delta_f > \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] &\leq \mathbb{E}_{u,\pi}[\Delta_f]\left(\frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right)^{-1} \\
&\underset{\text{by Eq. (17)}}{\leq} \varepsilon_f \frac{4}{|p_1 - p_2|}\max_u \frac{\delta}{\gamma}
\end{aligned} \tag{23}$$

$$\begin{aligned}
\mathbb{P}\left[\Delta_{\frac{\alpha}{\beta}} > \frac{1}{2\nu} \mid \Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] &\leq \mathbb{E}_{u,\pi}\left[\Delta_{\frac{\alpha}{\beta}} \mid \Delta_f \leq \frac{\gamma}{\delta}\frac{|p_1 - p_2|}{4}\right] \\
&\underset{\text{by Eq. (19)}}{\leq} \frac{\varepsilon_f}{|p_1 - p_2|}\max_u \frac{\beta\delta}{\alpha\gamma} \\
&\leq \frac{\varepsilon_f}{|p_1 - p_2|}\nu\max_u \frac{\delta}{\gamma}
\end{aligned} \tag{24}$$

Plugging back equations Eq. (21), Eq. (22),Eq. (23),Eq. (24) into equation Eq. (20), we obtain bounds for each term:

$$\mathbb{E}_{u,\pi}[\Delta_{\lambda^*}] = \underbrace{\mathbb{E}_{u,\pi}[\Delta_{\lambda^*} \mid A]\mathbb{P}[A]}_{\text{by Eq. (21)}} + \underbrace{\mathbb{E}_{u,\pi}\left[\Delta_{\lambda^*} \mid \bar{A}\right]}_{\text{by Eq. (22)}} \underbrace{\mathbb{P}\left[\bar{A}\right]}_{\text{by Eq. (23),Eq. (24)}} \tag{25}$$

we obtain:

$$\mathbb{E}_{u,\pi}[\Delta_{\lambda^*}] \leq \frac{\varepsilon_f}{|p_1 - p_2|}\left(\frac{\nu^3}{2} + \left(\nu + \frac{\nu^2}{4}\right)\mu\right) = \varepsilon_{\lambda^*}$$

To obtain the regret bound on the empirical rates, we apply assumptions (ii), (iii) once again to bound the expected difference between $\lambda^*(p)$ and $\frac{1}{T}|S_u(p;T)|$, and apply the triangle inequality:

$$\mathbb{E}_{u,\pi}\left[\left|\tfrac{1}{T}|S_u(p^*;T)| - \tfrac{1}{T}|S_u(\hat{p};T)|\right|\right] \leq \varepsilon_{\lambda^*} + 2(\varepsilon_{\text{dev}} + \varepsilon_{\text{LV}})$$

Note that $\frac{\nu}{|p_1-p_2|} > 1$, as $\frac{\beta}{\alpha} \geq 1$ since all the users are assumed to remain engaged in the long term, and $|p_1 - p_2| \leq 1$ as $p_1, p_2 \in [0,1]$. Therefore, the function $\eta_{\text{TPP}} = g(\mu,\nu) = \left(\frac{\nu^3}{2} + \left(\nu + \frac{\nu^2}{4}\right)\eta + 2\nu\right)$ satisfies:

$$\mathbb{E}_{u,\pi}\left[\left|\tfrac{1}{T}|S_u(p^*;T)| - \tfrac{1}{T}|S_u(\hat{p};T)|\right|\right] \leq \frac{\eta_{\text{TPP}}}{|p_1 - p_2|}(\varepsilon_{\text{pred}} + \varepsilon_{\text{dev}} + \varepsilon_{\text{LV}})$$

□

B Experimental details

B.1 Data

We base our experimental environment on the MovieLens 1M dataset, which is a standard benchmark dataset used widely in recommendation system research (Harper & Konstan, 2015). The dataset includes 1,000,209 ratings provided by 6,040 users and for 3,540 movies. Rating are in the range $\{1,\ldots,5\}$, and all users in the dataset have at least 20 reported ratings. The dataset is publicly available at: https://grouplens.org/datasets/movielens/1m/.

Data partitioning. To learn latent user and item features, 30% of all ratings were drawn at random. Stratified sampling was applied to ensure that all users and items were covered, and so that each users have roughly the same proportion of ratings used for this step. These ratings were only used only for learning a CF model, and were discarded afterwards. The remaining 70% data points were used for training and testing. For these, we first randomly sampled 1,000 users to form the test set. Then, the remaining users were partitioned into the main train set $\mathcal{S}$, which included 70% (≈3,528) of these users, and the experimental treatment sets $\mathcal{S}^{(j)}$, each including 10% (≈504) users for $N = 3$. This procedure was repeated 10 times, and we report average results and standard errors.

B.2 Implementation details

- **Hardware**: All experiments were run on a single laptop, with 16GB of RAM, M1 Pro processor, and with no GPU support.
- **Runtime**: A single run consisting the entire pipeline (data loading and partitioning, collaborative filtering, training classifiers, simulating dynamics, learning policies, measuring and comparing performance) takes roughly 17 minutes. The main bottleneck is the discrete LV simulation, taking roughly 70% of runtime to compute, mostly due to bookkeeping necessary for the non-stationary baselines. Simulation code was optimized using the NUMBA jit compiler, which improves runtime.
- **Optimization packages**:
  - **Collaborative filtering (CF)**: We use the SURPRISE package (Hug, 2020), which includes an implementation of the SVD algorithm for CF. All parameters were set to default values.
  - **Regression**: We use the SCIKIT-LEARN implementation of linear regression for predicting long-term engagement from user features (i.e the prediction models $f_j(u)$ in Eq. (8)). All parameters were set to default values.
  - **Non-Negative Least Squares (NNLS)**: We use the SCIPY.OPTIMIZE implementation of NNLS. The algorithm was used with its default parameters.
- **Code**: Code for reproducing all of our figures and experiments is available in the following anonymous repository: https://github.com/lvml-iclr-2023/lvml.

B.3 Other baselines

- **Safety**: In each step of the TPP simulation, look k step back, and calculate the empirical rate $\tilde{\lambda}_i = \frac{k}{t_i - t_{i-k}}$. If this rate exceeds the threshold $\tilde{\lambda}_i > \tau$, the policy enters a 'cool-down' policy

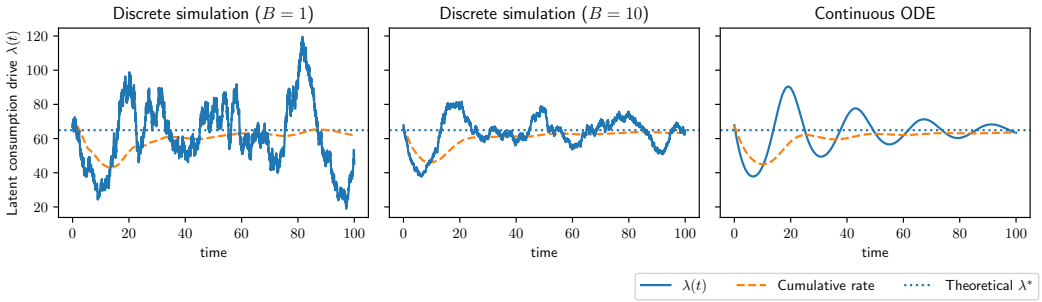

Figure 5: Example discrete sequence $S_u \sim \text{TPP}(u; T)$, compared to continuous LV dynamics. Our TPP produces discrete sequences that are qualitatively different from their continuous-time analogs (blue lines), Nonetheless, it captures the general properties of our proposed behavioral model: note how cumulative averaging behavior (orange dashes) exhibits 'habit formation', which our equilibrium approach targets (blue dots). For the same initial conditions $\lambda(0), q(0)$, the figure shows how varying the number of recommended items per step (B) 'smooths' the discrete behavior (left: $B = 1$, center: $B = 10$). For fixed $\beta_u(t) = \beta_u$, when $B \to \infty$, and when $\Delta t \to 0$, TPP sequences approach a continuous LV trajectory; in general, and particularly when $\beta_u(t)$ varies by step and per recommended items—this is not the case.

state, serving only forced breaks until the next time period. In our experiments, we used thresholds $\tau \in \{14, 16\}$, $k = 10$ look-behind steps, and defined the cool-down period as 0.5 time units.

- **Oracle**: To estimate the effect of perfect predictions, we implement an oracle predictor $f_p^{\text{oracle}}(u)$ which has access to the latent user parameters. For a given u and for each p, the predictor outputs the infinite-horizon LV equilibrium for u, namely $f_p^{\text{oracle}}(u) = \lambda^*(p; \tilde{\theta}_u)$. We define $\tilde{\theta}_u = (\alpha_u, \tilde{\beta}_u, \gamma_u, \delta_u)$, where $\alpha_u, \gamma_u, \delta_u$ are the unobserved parameters for the given user, and $\tilde{\beta}_u$ is the expected value of β_{ux} induced by the distribution over recommended items x induced by the recommendation policy ψ. We view $\tilde{\theta}_u$ as a useful proxy for the otherwise unattainable $\bar{\theta}_u$.

B.4 Hyperparameters

- **Collaborative filtering**: We used $d = 8$ latent factors and enabled bias terms, which ensured performance is close to the benchmark of $\text{RMSE} = 0.873$ reported in the SURPRISE documentation. We used the vanilla SVD solver, with all hyper-parameters set to their default values.
- **Recommendation policy**: Softmax temperature was set to 0.5.
- **Prediction**: We trained regressors $f(u)$ on input feature vectors $u \in \mathbb{R}^{d+2}$ consisting of: (i) SVD latent user factors, (ii) SVD user bias terms, (iii) an additional feature consisting of the average predicted ratings for unseen items (exponentiated and normalized), which we found to slightly improve predictive performance. We chose to focus on linear models since the treatment datasets are relatively small (each $|\mathcal{S}^{(j)}| \approx 500$), and since other model classes (including boosted trees and MLPs) did not perform significantly better.
- **Discrete TPP**: Interaction sequences for each user were generated according to an LV discretization scheme, described in detail in the next section. Latent sates were initialized randomly with relative uniform noise around the theoretical LV equilibrium point $(\lambda_0, q_0) = ((1 + \xi_\lambda)\lambda^*, (1 + \xi_q)q^*)$, where $\xi_\lambda, \xi_q \sim \text{Uniform}(-0.1, 0.1)$. Latent states were updated each $B = 10$ recommendations to stabilize noise (see Figure 5). When x is recommended to u at time t, latent states and Δt are set according to $\beta_u(t)$, which depends on ratings r_{ux} (true or mixed with predictions $u^\top x$ via κ). Specifically, we use $\beta_u(t) = r_{ux}^2/5 \in \{0.2, 0.8, 1.8, 3.2, 5\}$, which is convex, to accentuate the role of low ratings since they are underrepresented in the data. For $B \geq 1$, we take the effective $\beta_u(t)$ to be the average over the B items recommended in that step. We set $\alpha = 1.3$, and chose $\gamma = 0.2, \delta = 0.01$ (which together determine scale) so that typical values for engagement rate $\frac{1}{T}|S_u|$ are on the order of ≈ 10 for the chosen $T = 100$.

B.5 DISCRETE TPP

The TPP we use for simulating user behavior is based on a discretization of the LV system described in Eq. (2), based on the forward Euler method with variable step sizes.

Each user is associated with discrete latent states λ_i, q_i, and parameters $\alpha_u, \gamma_u, \delta_u$. Initial states λ_0, q_0 are set randomly. At each step, and in time t_i, the system recommends $x_i = x(t_i)$, which triggers updates in latent states, and determines the next time of interaction t_{i+1}. As noted, these update depend on item-specific parameters β_{u,x_i}.

Under stationary policy $\pi(p)$, the system recommends an item with probability $(1-p)$, and suggests a break with probability p. The simulator considers B recommendation opportunities at each step. For each $k \in \{1, \ldots, B\}$, denote by $I_k \in \{0, 1\}$ the break indicator, equal to 0 when a break is recommended at the k-th slot in the batch. Denote by $x \sim \psi$ the item recommended by the underlying policy ψ, and by $\beta(x) = r_{ux}^2/5$ the corresponding LV hyperparameter as defined above. For a given horizon T, the TPP process generating S_u is described by Algorithm 1:

Algorithm 1: Discrete TPP for user u

Input: Break probability $p \in [0, 1]$
Stationary content recommendation probability ψ
Time horizon $T > 0$
Output: Interaction sequence $S_u \sim \mathrm{TPP}_{\pi(p)\circ\psi}(u; T)$
Initialize $i = 0$, $t_0 = 0$, $S_u = \{\}$;
while $t_i < T$ **do**
  foreach $k \in \{1, \ldots, B\}$ **do**
    $I_k \sim \mathrm{Bernoulli}(1-p)$;
    $x_k \sim \psi$;
    $\beta_k \longleftarrow \beta(r_{ux_k})$;
  end
  $\Delta t_i = \lambda_i^{-1}$;
  $\lambda_{i+1} \longleftarrow \lambda_i + \left(-\alpha + \frac{\sum_{k=1}^{B} I_k \beta_k}{B} q_i\right) \lambda_i \Delta t_i$;
  $q_{i+1} \longleftarrow q_i + \left(\gamma(1-q_i) - \frac{\sum_{k=1}^{B} I_k \delta}{B} \lambda_i\right) q_i \Delta t_i$;
  $t_{i+1} \longleftarrow t_i + \lambda_{i+1}^{-1}$;
  $S_u \longleftarrow S_u \cup \{(t_i, (x_1, \ldots, x_B), (I_1, \ldots, I_B))\}$;
  $i \longleftarrow i + 1$;
end