A Organization of the Appendix

Here, we first present the proofs of results discussed in Section 3 of the main paper. The detailed proofs of the main results, Theorems 3.1 and 3.4, are provided in Section B of the appendix. All the other proofs of the intermediary results is provided in Section C of the Appendix. Thereafter, in Section D of the appendix we present further results based on numerical experiments.

B Detailed Proofs of Theorem 3.1 and 3.4

B.1 Proof of Theorem 3.1

The total regret $\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p})$ can be written down as the sum of regrets over all the segments and time period:

$$\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}) = \sum_{t=1}^{T} \sum_{l=1}^{L} \mathcal{R}_{lt} \,.$$

Using Proposition 4.1 and Lemma 4.2, we can bound the regret as

$$\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}) = C_9 C_{10} \sum_{t=1}^{T} \sum_{l=1}^{L} n_{lt} \langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt} \rangle^2 + C_9 C_{10} \sum_{t=1}^{T} \sum_{l=1}^{L} n_{lt} \big(p_{lt} (b_{lt} - \hat{b}_{lt}) \big)^2 \,.$$

Taking a maximum on the constants, we can rather bound the sum of the two terms $\sum_{t=1}^{T} \sum_{l=1}^{L} n_{lt} \langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt} \rangle^2$ and $\sum_{t=1}^{T} \sum_{l=1}^{L} n_{lt} \big(p_{lt} (b_{lt} - \hat{b}_{lt}) \big)^2$ to get a final bound on the regret.

We do the analysis for a fixed segment l first and then combine the regret across all the segments.

Lemma B.1. *Consider model (12), true parameters $\boldsymbol{m}_{lt}$, b_{lt} and the output $\hat{\boldsymbol{m}}_{lt}$, $\hat{b}_{lt}$ from our*

PSGD pricing policy, the following holds with probability at least $1 - T^{-2}$

$$\sum_{t=1}^{T} n_{lt}\Big(\langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt}\rangle^2 + \big(p_{lt}(b_{lt} - \hat{b}_{lt})\big)^2\Big)$$
$$\leq C_1 \sum_{t=1}^{T} \frac{1}{\eta_t}\|\boldsymbol{m}_{l,t+1} - \boldsymbol{m}_{lt}\|_2 + C_2 \sum_{t=1}^{T} \frac{1}{\eta_t}|b_{l,t+1} - b_{lt}| + C_3 \sum_{t=1}^{T} \eta_t n_{lt}^2 + \frac{C_4}{\eta_{T+1}} + \mathcal{O}(\log T)\,.$$

With this lemma we have with probability at least $1 - L/T^2$,

$$\sum_{t=1}^{T}\sum_{l=1}^{L} n_{lt}\Big(\langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt}\rangle^2 + \big(p_{lt}(b_{lt} - \hat{b}_{lt})\big)^2\Big)$$
$$\leq C_1 \sum_{t=1}^{T}\sum_{l=1}^{L} \frac{1}{\eta_t}\|\boldsymbol{m}_{l,t+1} - \boldsymbol{m}_{lt}\|_2 + C_2 \sum_{t=1}^{T}\sum_{l=1}^{L} \frac{1}{\eta_t}|b_{l,t+1} - b_{lt}| + C_3 \sum_{t=1}^{T}\sum_{l=1}^{L} \eta_t n_{lt}^2 + \frac{C_4 L}{\eta_{T+1}} + \mathcal{O}(\log T)\,.$$

Define the RHS as I.

Consider $\mathcal{G}$ to be the probabilistic event that the above is true, then $\mathbb{P}(\mathcal{G}^C) = L/T^2$.

Also, since the maximum price is M and we set a positive price, hence the maximum revenue lost on the event $\mathcal{G}^C$ is $\sum_{t=1}^{T} n_t M$. Assuming that $N_T = \max_{t \leq T} n_T$, we have the maximum regret in the event $\mathcal{G}^C$ is TMN_T.

The total regret is thus,

$$\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}) = \mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}|\mathcal{G}) + \mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}|\mathcal{G}^C) \leq I\mathbb{P}(\mathcal{G}) + TN_T\mathbb{P}(\mathcal{G}^C) \leq A + \frac{MLN_T}{T}\,.$$

Since the last term is $\mathcal{O}(1/T)$, we have the required terms of the regret bound.

B.1.1 Proof of Lemma B.1

Let $\boldsymbol{\psi}_{lt} = (\boldsymbol{m}_{lt}, b_{lt})$ be the combined parameter space and $\boldsymbol{Q}_{lt} = (\boldsymbol{x}_{lt}, p_{lt})$ be the covariates and the price posted.

By Taylor expansion of the loss function we get, for some $\tilde{\boldsymbol{\psi}}_{l,t}$ between $\hat{\boldsymbol{\psi}}_{l,t}$ and $\boldsymbol{\psi}_{lt}$,

$$\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) - \mathcal{L}_{lt}(\boldsymbol{\psi}_{lt}) = \langle \nabla\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle - \frac{1}{2}\langle \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, \nabla^2\mathcal{L}_{lt}(\tilde{\boldsymbol{\psi}}_{l,t})(\hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt})\rangle\,. \tag{30}$$

Simplifying the loss function in terms of $\boldsymbol{\psi}_{lt}$ and $\boldsymbol{Q}_{lt}$ gives us

$$\mathcal{L}_{lt}(\boldsymbol{\psi}) = -\left(y_{lt} \log \Phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi}) + \tilde{y}_{lt} \log \Phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi})\right),$$

where $\tilde{y}_{lt} = n_{lt} - y_{lt}$. The second derivative of the loss function can thus be computed as

$$\nabla^2 \mathcal{L}_{lt}(\boldsymbol{\psi}) = -\left(y_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}} + \tilde{y}_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}} \right).$$

Let $c_{\mathcal{L}} = \min\left\{-\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}}, -\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\right\}$. Based on our assumptions, $\boldsymbol{Q}_{lt}$ and $\boldsymbol{\psi}$ are bounded and so there exists c such that $|\boldsymbol{Q}_{lt}\boldsymbol{\psi}| < c$. Since Φ is log-concave hence the second derivative is negative and $-\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta) > 0$. Particularly since the second derivative only approaches 0 when ζ goes to ∞ or $-\infty$, hence on the bounded set with $|\zeta| < c$, second derivative is bounded away from zero implying $c_{\mathcal{L}} > 0$.

Then,

$$\begin{aligned}
\nabla^2 \mathcal{L}_{lt}(\boldsymbol{\psi}) &= -\left(y_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}} + \tilde{y}_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}} \right) \\
&= \left(y_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \left(-\frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\right) + \tilde{y}_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \left(-\frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\right) \right) \\
&\geq (y_{lt} + \tilde{y}_{lt})\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T \min\left\{-\frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}}, -\frac{\partial}{\partial \zeta^2} \log \Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\right\} \\
&= n_{lt}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T c_{\mathcal{L}}\,.
\end{aligned}$$

Where the last equality follows since $y_{lt} + \tilde{y}_{lt} = n_{lt}$.

Using this in (30)

$$\begin{aligned}
\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) - \mathcal{L}_{lt}(\boldsymbol{\psi}_t) &\leq \langle \nabla\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle - \frac{1}{2}\langle \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, n_{lt}c_{\mathcal{L}}\boldsymbol{Q}_{lt}\boldsymbol{Q}_{lt}^T(\hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt})\rangle \\
&= \langle \nabla\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle - \frac{n_{lt}c_{\mathcal{L}}}{2}\langle \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, \boldsymbol{Q}_{lt}\rangle^2 \\
&= \langle \nabla\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{l,t+1} - \boldsymbol{\psi}_{lt}\rangle + \langle \nabla\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}\rangle - \frac{n_{lt}c_{\mathcal{L}}}{2}\langle \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, \boldsymbol{Q}_{lt}\rangle^2\,.
\end{aligned} \tag{31}$$

Our update rules in (19), gives us

$$\hat{b}_{l,t+1} = \Pi_{\Theta_b}(\hat{b}_{lt} - \eta_t \nabla \mathcal{L}^b_{lt})\,,$$
$$\hat{\boldsymbol{m}}_{l,t+1} = \Pi_{\Theta_{\boldsymbol{m}}}(\hat{\boldsymbol{m}}_{lt} - \eta_t \nabla \mathcal{L}^{\boldsymbol{m}}_{lt})\,.$$

The updates defined are common OMD updates and can be rewritten as

$$\hat{\boldsymbol{\psi}}_{l,t+1} = \arg\min_{\boldsymbol{\psi}} \langle \nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \boldsymbol{\psi}\rangle + \frac{1}{2\eta_t}\|\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_{lt}\|^2\,.$$

Since the above loss function is convex and $\hat{\boldsymbol{\psi}}_{l,t+1}$ is the minimizer, we get

$$\langle \boldsymbol{\psi} - \hat{\boldsymbol{\psi}}_{l,t+1}, \eta_t \nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) + \hat{\boldsymbol{\psi}}_{l,t+1} - \hat{\boldsymbol{\psi}}_{lt}\rangle \geq 0\,.$$

Putting $\boldsymbol{\psi} = \boldsymbol{\psi}_{lt}$ above, we get $\langle \hat{\boldsymbol{\psi}}_{l,t+1} - \boldsymbol{\psi}_{lt}, \eta_t \nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})\rangle \leq \langle \boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}, \hat{\boldsymbol{\psi}}_{l,t+1} - \hat{\boldsymbol{\psi}}_{lt}\rangle$.

Also, note that

$$\langle \boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}, \hat{\boldsymbol{\psi}}_{l,t+1} - \hat{\boldsymbol{\psi}}_{lt}\rangle = \frac{1}{2}\left(\|\boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{lt}\|^2 - \|\boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}\|^2 - \|\hat{\boldsymbol{\psi}}_{l,t+1} - \hat{\boldsymbol{\psi}}_{lt}\|^2\right)\,.$$

With the above two equations the first term in (31) is bounded as:

$$\langle \nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{l,t+1} - \boldsymbol{\psi}_{lt}\rangle \leq \frac{1}{2\eta_t}\left(\|\boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{lt}\|^2 - \|\boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}\|^2 - \|\hat{\boldsymbol{\psi}}_{l,t+1} - \hat{\boldsymbol{\psi}}_{lt}\|^2\right)\,.$$

Using the inequality $ab \leq (a^2 + b^2)/2$, the second term in (31) can be bounded as,

$$\langle \nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}), \hat{\boldsymbol{\psi}}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}\rangle \leq \frac{1}{2\eta_t}\|\hat{\boldsymbol{\psi}}_{lt} - \hat{\boldsymbol{\psi}}_{l,t+1}\|^2 + \frac{\eta_t}{2}\|\nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})\|^2\,. \tag{32}$$

Also, $\nabla \mathcal{L}_{lt}(\boldsymbol{\psi}) = -\left(y_{lt}\boldsymbol{Q}_{lt}\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}} - \tilde{y}_{lt}\boldsymbol{Q}_{lt}\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\right)$.

Let $C_{\mathcal{L}} = \max\{-\frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=\boldsymbol{Q}_{lt}\boldsymbol{\psi}}, \frac{\partial}{\partial\zeta^2}\log\Phi(\zeta)|_{\zeta=-\boldsymbol{Q}_{lt}\boldsymbol{\psi}}\}$ in the restricted space. Hence, $\|\nabla \mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})\|^2 \leq C^2_{\mathcal{L}} n^2_{lt}\|\boldsymbol{Q}_{lt}\|^2$.

Combining all the parts we have,

$$\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})-\mathcal{L}_{lt}(\boldsymbol{\psi}_t) \leq \frac{1}{2\eta_t}\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{lt}\|^2-\frac{1}{2\eta_t}\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2+\frac{\eta_t}{2}C_{\mathcal{L}}^2 n_{lt}^2\|\boldsymbol{Q}_{lt}\|^2-\frac{n_{lt}c_{\mathcal{L}}}{2}\langle\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt},\boldsymbol{Q}_{lt}\rangle^2 .$$

Adding and subtracting $\|\boldsymbol{\psi}_{l,t+1}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2$ to above we get

$$\begin{aligned}\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})-\mathcal{L}_{lt}(\boldsymbol{\psi}_t) \leq\ & \frac{1}{2\eta_t}\left(\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{lt}\|^2-\|\boldsymbol{\psi}_{l,t+1}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2\right)\\ &+\frac{1}{2\eta_t}\left(\|\boldsymbol{\psi}_{l,t+1}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2-\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2\right)\\ &+\frac{\eta_t}{2}C_{\mathcal{L}}^2 n_{lt}^2\|\boldsymbol{Q}_{lt}\|^2-\frac{n_{lt}c_{\mathcal{L}}}{2}\langle\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt},\boldsymbol{Q}_{lt}\rangle^2 . \end{aligned} \tag{33}$$

The second term can be simplified as

$$\|\boldsymbol{\psi}_{l,t+1}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2-\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{l,t+1}\|^2=\langle\boldsymbol{\psi}_{l,t+1}+\boldsymbol{\psi}_{lt}-2\hat{\boldsymbol{\psi}}_{l,t+1},\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\rangle \leq 4C_{\boldsymbol{\psi}}\|\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\|_2 ,$$

where $C_{\boldsymbol{\psi}}$ is $\max\|\boldsymbol{\psi}\|$ and $C_{\boldsymbol{\psi}} \leq 2C_b+2C_m$, since $\boldsymbol{\psi}=(\boldsymbol{m},b)$.

Summing both sides of (33) over $t=1,\ldots,T$, we get $\sum_{t=1}^T\left(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})-\mathcal{L}_{lt}(\boldsymbol{\psi}_t)\right)$ is bounded above by:

$$\begin{aligned}&\frac{\|\boldsymbol{\psi}_{l1}-\hat{\boldsymbol{\psi}}_{l1}\|^2}{2\eta_1}+\sum_{t=2}^T\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{lt}\|^2\left(\frac{1}{2\eta_{t+1}}-\frac{1}{2\eta_t}\right)+4C_{\boldsymbol{\psi}}\sum_{t=1}^T\frac{1}{2\eta_t}\|\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\|_2\\ &+\sum_{t=1}^T\frac{\eta_t}{2}C_{\mathcal{L}}^2 n_{lt}^2\|\boldsymbol{Q}_{lt}\|^2-\sum_{t=1}^T\frac{n_{lt}c_{\mathcal{L}}}{2}\langle\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt},\boldsymbol{Q}_{lt}\rangle^2 .\end{aligned}$$

Under the assumption that η_t are non-decreasing,

$$\frac{\|\boldsymbol{\psi}_{l1}-\hat{\boldsymbol{\psi}}_{l1}\|^2}{2\eta_1}+\sum_{t=2}^T\|\boldsymbol{\psi}_{lt}-\hat{\boldsymbol{\psi}}_{lt}\|^2\left(\frac{1}{2\eta_{t+1}}-\frac{1}{2\eta_t}\right) \leq \frac{4C_{\boldsymbol{\psi}}^2}{2\eta_1}+4C_{\boldsymbol{\psi}}^2\sum_{t=2}^T\left(\frac{1}{2\eta_{t+1}}-\frac{1}{2\eta_t}\right)=\frac{4C_{\boldsymbol{\psi}}^2}{2\eta_{T+1}} .$$

Hence, we finally have

$$\sum_{t=1}^{T}\Big(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})-\mathcal{L}_{lt}(\boldsymbol{\psi}_t)\Big)\le \frac{4C_{\boldsymbol{\psi}}^2}{2\eta_{T+1}}+4C_{\boldsymbol{\psi}}\sum_{t=1}^{T}\frac{1}{2\eta_t}\|\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\|_2 + \sum_{t=1}^{T}\frac{\eta_t}{2}C_{\mathcal{L}}^2 n_{lt}^2\|\boldsymbol{Q}_{lt}\|^2-\sum_{t=1}^{T}\frac{n_{lt}c_{\mathcal{L}}}{2}\langle\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt},\boldsymbol{Q}_{lt}\rangle^2\,.$$

Define

$$A:=\frac{4C_{\boldsymbol{\psi}}^2}{2\eta_{T+1}}+4C_{\boldsymbol{\psi}}\sum_{t=1}^{T}\frac{1}{2\eta_t}\|\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\|_2+\sum_{t=1}^{T}\frac{\eta_t}{2}C_{\mathcal{L}}^2 n_{lt}^2\|\boldsymbol{Q}_{lt}\|^2. \tag{34}$$

Since, $\|\boldsymbol{\psi}_{l,t+1}-\boldsymbol{\psi}_{lt}\|_2\le 2(\|\boldsymbol{m}_{l,t+1}-\boldsymbol{m}_{lt}\|_2+|b_{l,t+1}-b_{lt}|)$ and $\|\boldsymbol{Q}_{lt}\|^2$ is bounded, we can simplify A as

$$A:=\tilde{C}_1\sum_{t=1}^{T}\frac{1}{\eta_t}\|\boldsymbol{m}_{l,t+1}-\boldsymbol{m}_{lt}\|_2+\tilde{C}_2\sum_{t=1}^{T}\frac{1}{\eta_t}|b_{l,t+1}-b_{lt}|+\tilde{C}_3\sum_{t=1}^{T}\eta_t n_{lt}^2+\frac{\tilde{C}_4}{\eta_{T+1}}\,.$$

Note that in order to prove the lemma we need to show a bound on $\sum_{t=1}^{T} n_{lt}\Big(\langle\boldsymbol{x}_{lt},\boldsymbol{m}_{lt}-\hat{\boldsymbol{m}}_{lt}\rangle^2+\big(p_{lt}(b_{lt}-\hat{b}_{lt})\big)^2\Big)$ which is same as showing a bound on $\sum_{t=1}^{T} n_{lt}\langle\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt},\boldsymbol{Q}_{lt}\rangle^2$, since $\boldsymbol{\psi}_{lt}=(\boldsymbol{m}_{lt},b_{lt})$ and $\boldsymbol{Q}_{lt}=(\boldsymbol{x}_{lt},p_{lt})$.

We next provide a lower bound on the cumulative difference $\sum_{t=1}^{T}\Big(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt})-\mathcal{L}_{lt}(\boldsymbol{\psi}_t)\Big)$. Write

$$\mathcal{L}_{ltk}(\boldsymbol{\psi}_{lt})-\mathcal{L}_{ltk}(\hat{\boldsymbol{\psi}}_{lt})\le\langle\nabla\mathcal{L}_{ltk}(\boldsymbol{\psi}_{lt}),\hat{\boldsymbol{\psi}}_{lt}-\boldsymbol{\psi}_{lt}\rangle:=D_{tk}\,, \tag{35}$$

using convexity of the loss $\mathcal{L}_{ltk}$. We also have

$$\begin{aligned}\nabla\mathcal{L}_{ltk}(\boldsymbol{\psi})&=-(y_{ltk}\frac{\partial}{\partial\boldsymbol{\psi}}\log\Phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})-(1-y_{ltk})\frac{\partial}{\partial\boldsymbol{\psi}}\log\Phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi}))\\&=\boldsymbol{Q}_{lt}\left(-y_{ltk}\frac{\phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})}{\Phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})}+(1-y_{ltk})\frac{\phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi})}{\Phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi})}\right)\,.\end{aligned}$$

Let $\mathcal{F}_t$ be the σ-field generated by the noise till time t. Then, since $\hat{\boldsymbol{\psi}}_{lt}$ only depends on noise till time t, $\mathbb{E}[D_{tk}|\mathcal{F}_{t-1}] = \langle \mathbb{E}[\nabla \mathcal{L}_{ltk}|\mathcal{F}_{t-1}], \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle$. In addition, $E[\nabla \mathcal{L}_{ltk}|\mathcal{F}_{t-1}] = 0$ using the fact that $\mathbb{P}(Y_{ltk} = 1) = \Phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})$ and $\mathbb{P}(Y_{ltk} = 0) = \Phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi}))$. Therefore, the partial sums of D_{tk} is a martingale with respect to the filtration $\mathcal{F}_t$.

Also, as described above $\left(-y_{ltk}\frac{\phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})}{\Phi(\boldsymbol{Q}_{lt}\boldsymbol{\psi})} + (1-y_{ltk})\frac{\phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi})}{\Phi(-\boldsymbol{Q}_{lt}\boldsymbol{\psi})}\right)$ is bounded above with $C_{\mathcal{L}}$. Hence $|D_{tk}| \leq \beta_t := C_{\mathcal{L}}|\langle \boldsymbol{Q}_{lt}, \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle|$. Using convexity of $e^{\lambda z}$, for any $\lambda \in \mathbb{R}$ we have

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda D_{tk}} \mid \mathcal{F}_{t-1}\right] &\leq \mathbb{E}\left[\frac{\beta_t - D_{tk}}{2\beta_t}e^{-\lambda\beta_t} + \frac{\beta_t + D_{tk}}{2\beta_t}e^{\lambda\beta_t} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\frac{e^{-\lambda\beta_t} + e^{\lambda\beta_t}}{2}\right] + \mathbb{E}\left[D_{tk} \mid \mathcal{F}_{t-1}\right]\left(\frac{e^{-\lambda\beta_t} + e^{\lambda\beta_t}}{2\beta_t}\right) = \cosh\left(\lambda\beta_t\right) \leq e^{\lambda^2\beta_t^2/2}.
\end{aligned}$$

where $\beta_t = C_{\mathcal{L}}|\langle \boldsymbol{Q}_{lt}, \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}\rangle|$. We next use the following result from (Javanmard 2017, Proposition C.1).

Proposition B.2. *(Javanmard 2017, Proposition C.1) Consider a martingale difference sequence D_t adapted to a filtration $\mathcal{F}_t$, such that for any $\lambda \geq 0, \mathbb{E}\left[e^{\lambda D_t} \mid \mathcal{F}_{t-1}\right] \leq e^{\lambda^2\sigma_t^2/2}$. Then, for $D(T) = \sum_{t=1}^T D_t$, the following holds true:*

$$\mathbb{P}(D(T) \geq \xi) \leq e^{-\xi^2/\left(2\sum_{t=1}^T \sigma_t^2\right)}.$$

We apply the above theorem with $D(T) = \sum_{t=1}^T \sum_{k=1}^{n_{lt}} D_{tk}$. Invoking (35), this gives us,

$$\mathbb{P}\left(\sum_{t=1}^T \left(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) - \mathcal{L}_{lt}(\boldsymbol{\psi}_t)\right) \leq -2C_{\mathcal{L}}\sqrt{\log T}\left\{\sum_{t=1}^T n_{lt}\left\langle \boldsymbol{Q}_{lt}, \boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{lt}\right\rangle^2\right\}^{1/2}\right) \leq \frac{1}{T^2}.$$

Hence with probability at least $1 - 1/T^2$,

$$\sum_{t=1}^T \left(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) - \mathcal{L}_{lt}(\boldsymbol{\psi}_t)\right) \geq -2C_{\mathcal{L}}\sqrt{\log T}\left\{\sum_{t=1}^T n_{lt}\left\langle \boldsymbol{Q}_{lt}, \boldsymbol{\psi}_{lt} - \hat{\boldsymbol{\psi}}_{lt}\right\rangle^2\right\}^{1/2}.$$

Let $B = \sum_{t=1}^T n_{lt}\langle \hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, \boldsymbol{Q}_{lt}\rangle^2$, then with the complete analysis till now we have

$$-2C_{\mathcal{L}}\sqrt{B\log T} \le \sum_{t=1}^{T}\Big(\mathcal{L}_{lt}(\hat{\boldsymbol{\psi}}_{lt}) - \mathcal{L}_{lt}(\boldsymbol{\psi}_t)\Big) \le A - \frac{c_{\mathcal{L}}}{2}B\,.$$

Hence, $B - (4C_{\mathcal{L}}/c_{\mathcal{L}})\sqrt{B\log T} \le (2/c_{\mathcal{L}})A$. Consider two cases:

Case 1: $\sqrt{B\log T} \le (c_{\mathcal{L}}/8C_{\mathcal{L}})B$, then $B \le 4A/c_{\mathcal{L}}$.

Case 2: $\sqrt{B\log T} \ge (c_{\mathcal{L}}/8C_{\mathcal{L}})B$, then $B \le (8C_{\mathcal{L}}/c_{\mathcal{L}})^2\log T$.

Combining the two cases, we have

$$B \le \frac{4A}{c_{\mathcal{L}}} + \mathcal{O}(\log T)\,.$$

Substituting for

$$B = \sum_{t=1}^{T} n_{lt}\langle\hat{\boldsymbol{\psi}}_{lt} - \boldsymbol{\psi}_{lt}, \boldsymbol{Q}_{lt}\rangle^2 = \sum_{t=1}^{T} n_{lt}\left(\langle\boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt}\rangle^2 + p_{lt}^2(b_{lt} - \hat{b}_{lt})^2\right)\,,$$

and A from (34) we obtain the desired result.

B.2 Proof of Theorem 3.4

Recall the varaince of segment l in the utility model given by $V_{lt}^2 = \|(\boldsymbol{I} - \rho_t\boldsymbol{W})^{-1}\boldsymbol{e}_l\|^2\tau^2 + \sigma^2$. We assume that we have a known fixed ρ, the auto-correlation parameter, and so the variances do not change over time.

Indicate the variances of segments by $V_1, V_2, \cdots, V_L$. Our utility model is thus

$$\tilde{U}_{ltk} = \frac{\beta_t}{V_l}p_{lt} + \boldsymbol{x}_{lt}'\frac{\boldsymbol{\mu}_t}{V_l} + Z_{ltk}\,.$$

Without loss of generality, assume that $\boldsymbol{x}_{lt}$ is of dimension one. We would give a small variation that would work for any dimension as well. Let $v_1, v_2, \cdots, v_L$ be the inverse of the fixed variances. The model is thus,

$$\tilde{U}_{ltk} = \beta_t v_l\, p_{lt} + x_{lt}\mu_t v_l + Z_{ltk}\,.$$

Assume that $-\beta_t = \mu_t = \gamma$, i.e. the parameters do not change over time and are negative of each other. In this setup $U^0_{ltk} = v_l\gamma(x_{lt} - p_{lt})$, the noiseless utility. Here setting $p_{lt} = x_{lt}$ would be uninformative since we would just observe noise, and we cannot get any information about the unknown parameter γ. In addition, in our model a price p^*_{lt} is optimum if it satisfies

$$p^*_{lt} = -\frac{1}{\beta_t v_l}\frac{\Phi(\beta_t v_l\, p^*_{lt} + x_{lt}\mu_t v_l)}{\phi(\beta_t v_l\, p^*_{lt} + x_{lt}\mu_t v_l)}\,.$$

Under the assumption that $-\beta_t = \mu_t = \gamma$, this reduces to

$$p^*_{lt} = \frac{1}{\gamma v_l}\frac{\Phi(v_l\gamma(x_{lt} - p^*_{lt}))}{\phi(v_l\gamma(x_{lt} - p^*_{lt}))}\,.$$

Therefore, for $\gamma_0 := (v_l x_{lt})^{-1}\Phi(0)/\phi(0)$, the uninformative price is optimal prices, i.e., $p^*_{lt}(\gamma_0) = x_{lt}$.

Note that if x_{lt} was of higher dimension, we could set $\boldsymbol{x}_{lt} = (a/d, a/d, a/d, \cdots, a/d)$ with d the dimension of $\boldsymbol{x}_{lt}$ and set $\boldsymbol{\mu}_t = (\gamma, \gamma, \gamma, \cdots, \gamma)$ to get the exactly same result: $p^*_{lt}(\gamma_0) = a$ for $\gamma_0 = (v_l a)^{-1}\Phi(0)/\phi(0)$ is an uniformative price.

Now that we know the existence of a setting where the uninformative prices are optimal prices, we can show that the regret is at least of the order of $\sqrt{T}$.

We construct a problem class $(\Gamma, \{\mathcal{P}_{lt}\})$, for $l = 1, \ldots, L$, $t = 1, \ldots, T$ as follows. Recall γ_0 the parameter for which the optimal price is uninformative. We use the shorthand $r_{lt}(p, \gamma)$ to denote the expected revenue obtained from a typical customer from segment l at time t, if the model parameter is γ. Therefore, recalling our utility model $U_{ltk} = v_l\gamma(x_{lt} - p_{lt}) + Z_{ltk}$, we have $r_{lt}(p, \gamma) = p(\Phi(v_l\gamma(x_{lt} - p)))$. By optimality of $p^*_{lt}(\gamma_0)$, we have $r''_{lt}(p^*_{lt}(\gamma_0), \gamma_0) < -2c$ for some constant $c > 0$, and by continuity of r''_{lt} we can find a neighborhood $\mathcal{P}_{lt}$ around $p^*_{lt}(\gamma_0)$ such that $r''_{lt}(p, \gamma_0) < -c$ for all $p \in \mathcal{P}_{lt}$. We next consider the mapping $\gamma \mapsto p^*_{lt}(\gamma)$. By continuity of this mapping, we can find a small enough neighborhood Γ_{lt} around γ_0 such that the optimal prices $p^*_{lt}(\gamma) \in \mathcal{P}_{lt}$ for all $\gamma \in \Gamma_{lt}$. Finally, we take $\Gamma := \cap_{t=1}^T \cap_{l=1}^L \Gamma_{lt}$. Note that Γ is non-empty because $\gamma_0 \in \Gamma$. Furthermore, by our construction we have the following properties for the problem class $(\Gamma, \{\mathcal{P}\}_{lt})$, for $l = 1, \ldots, L$ and $t = 1, \ldots, T$:

- For all $\gamma \in \Gamma$, we have $p^*_{lt}(\gamma) \in \mathcal{P}_{lt}$.

- For all prices $p \in \mathcal{P}_{lt}$, we have $r''_{lt}(p, \gamma_0) < -c$.

For any pricing policy π and a parameter $\gamma \in \Gamma$, let $f_t^{\pi,\gamma} : \{0,1\}^{N_t} \to [0,1]$ be the probability distribution function for all the consumers purchase responses $\boldsymbol{Y} = (Y_{ljk}, \ell = 1, \dots, L, j = 1, \dots, t, k = 1, \dots, n_{lj})$ until time t. Here, $N_t = \sum_{j=1}^t \sum_{l=1}^L n_{lj}$, under policy π and model parameter γ. The pricing policy uses all the sales data till time $t-1$ to give a price p^*_{lt}. We use $\boldsymbol{y}_t \in \{0,1\}^{N_t}$ to denote all sales data till time t. So, if the pricing policy gives the prices $p_{lt} := \pi(\boldsymbol{y}_{t-1})$ for all the time periods, then

$$f_t^{\pi,\gamma}(\boldsymbol{y}_t) = \prod_{j=1}^{t} \prod_{l=1}^{L} \prod_{k=1}^{n_{lj}} q_{lj}(p_{lj}, \gamma)^{y_{ljk}} (1 - q_{lj}(p_{lj}, \gamma))^{1 - y_{ljk}} ,$$

where $q_{lj}(p_{lj}, \gamma) = \Phi(v_l \gamma (x_{lj} - p_{lj}))$. We next want to show that for γ_0, the parameter for which the uninformative price is optimal, any policy incurs a large regret if it tries to learn γ_0. Formally, we aim to show that

$$\mathcal{R}_t^\pi(\gamma_0) \geq C \frac{1}{(\gamma_0 - \gamma)^2} \mathsf{KL}(f_t^{\pi,\gamma_0}, f_t^{\pi,\gamma}) .$$

We employ the chain rule for KL divergence (Cover & Thomas 1991),

$$\begin{aligned}
&\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) = \sum_{s=1}^{t} \mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma} | \mathbf{Y}_{s-1}\right) \\
&= \sum_{s=1}^{t} \sum_{\mathbf{y}_s \in \{0,1\}^{N_s}} f_s^{\pi,\gamma_0}(\mathbf{y}_s) \log \left(\frac{f_s^{\pi,\gamma_0}(y_{lsk} \mid \mathbf{y}_{s-1})}{f_s^{\pi,\gamma}(y_{lsk} \mid \mathbf{y}_{s-1})} \right) \\
&= \sum_{s=1}^{t} \sum_{\mathbf{y}_{s-1} \in \{0,1\}^{N_{s-1}}} f_{s-1}^{\pi,\gamma_0}(\mathbf{y}_{s-1}) \sum_{l=1}^{L} \sum_{k=1}^{n_{ls}} \sum_{y_{lsk} \in \{0,1\}} f_s^{\pi,\gamma_0}(y_{lsk} \mid \mathbf{y}_{s-1}) \log \left(\frac{f_s^{\pi,\gamma_0}(y_{lsk} \mid \mathbf{y}_{s-1})}{f_s^{\pi,\gamma}(y_{lsk} \mid \mathbf{y}_{s-1})} \right) \\
&= \sum_{s=1}^{t} \sum_{\mathbf{y}_{s-1} \in \{0,1\}^{N_{s-1}}} f_{s-1}^{\pi,\gamma_0}(\mathbf{y}_{s-1}) \sum_{l=1}^{L} \sum_{k=1}^{n_{ls}} \mathsf{KL}\Big(f_s^{\pi,\gamma_0}(y_{lsk} \mid \mathbf{y}_{s-1}) ; f_s^{\pi,\gamma}(y_{lsk} \mid \mathbf{y}_{s-1}) \Big) .
\end{aligned}$$

Based on the definition of $f_t^{\pi,\gamma}$, $f_s^{\pi,\gamma_0}(y_{lsk})$ is distributed as Bernoulli $q_{ls}(p_{ls}, \gamma_0)$ and $f_s^{\pi,\gamma}(y_{lsk})$

is distributed as Bernoulli $q_{ls}(p_{ls}, \gamma)$. Using the fact that for Bernoulli random variables $B_1 \sim \mathsf{Bern}(q_1), B_2 \sim \mathsf{Bern}(q_2)$, we have $\mathsf{KL}(B_1, B_2) \le \frac{(q_1-q_2)^2}{q_2(1-q_2)}$, we get

$$\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) \le \sum_{s=1}^{t} \sum_{\mathbf{y}_{s-1}\in\{0,1\}^{N_{s-1}}} f_{s-1}^{\pi,\gamma_0}\left(\mathbf{y}_{s-1}\right) \sum_{l=1}^{L} \sum_{k=1}^{n_{ls}} \frac{(q_{ls}(p_{ls},\gamma_0) - q_{ls}(p_{ls},\gamma))^2}{q_{ls}(p_{ls},\gamma)(1-q_{ls}(p_{ls},\gamma))}\,. \tag{36}$$

Since the prices and the parameters are bounded, and q_{lt} is the normal distribution function, q_{lt} is bounded away from zero. Hence, there exists constant C such that $q_{ls}(1-q_{ls}) \ge C$.

Also, $q_{ls}(p^*_{ls}, \gamma) = \Phi(v_l\gamma(x_{ls} - p^*_{ls}))$ and $q_{ls}(p^*_{ls}, \gamma_0) = \Phi(v_l\gamma_0(x_{ls} - p^*_{ls}))$. Since we are working on a bounded set, the distribution function Φ is Lipschitz as well. Hence,

$$\begin{aligned}
q_{ls}(p_{ls},\gamma_0) - q_{ls}(p_{ls},\gamma) &= \Phi(v_l\gamma_0(x_{ls}-p_{ls})) - \Phi(v_l\gamma(x_{ls}-p_{ls}))\\
&\le C(v_l\gamma_0(x_{ls}-p_{ls}) - v_l\gamma(x_{ls}-p_{ls}))\\
&= Cv_l(\gamma_0-\gamma)(x_{ls}-p_{ls})\\
&= Cv_l(\gamma_0-\gamma)(p^*_{ls}(\gamma_0)-p_{ls})\,,
\end{aligned}$$

where $p^*_{ls}(\gamma_0)$ is the optimal price for when the parameter is γ_0. Recall that by the definition of γ_0, the optimal price for γ_0 is x_{lt}. We thus have

$$(q_{ls}(p_{ls},\gamma_0) - q_{ls}(p_{ls},\gamma))^2 \le C(\gamma_0-\gamma)^2(p^*_{ls}(\gamma_0)-p_{ls})^2\,.$$

Using the above bound in (36), we get

$$\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) \le C(\gamma-\gamma_0)^2 \sum_{s=1}^{t}\sum_{l=1}^{L}\sum_{k=1}^{n_{ls}} \sum_{\mathbf{y}_{s-1}\in\{0,1\}^{N_{s-1}}} f_{s-1}^{\pi,\gamma_0}\left(\mathbf{y}_{s-1}\right)\left(p^*_{ls}(\gamma_0)-p_{ls}\right)^2.$$

The inner summation is indeed the expectation with respect to γ_0, by noting that p_{ls} is a measurable function of $\boldsymbol{y}_{s-1}$. Hence, we have

$$\begin{aligned}\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) &\le C(\gamma-\gamma_0)^2 \sum_{s=1}^{t}\sum_{l=1}^{L}\sum_{k=1}^{n_{ls}} \mathbb{E}_{\gamma_0}(p_{ls}^*(\gamma_0)-p_{ls})^2 \\ &= C(\gamma-\gamma_0)^2 \sum_{s=1}^{t}\sum_{l=1}^{L} n_{ls}\mathbb{E}_{\gamma_0}(p_{ls}^*(\gamma_0)-p_{ls})^2\,. \end{aligned} \tag{37}$$

By the construction of problem class $(\Gamma, \{\mathcal{P}_{lt}\})$, we have $r''_{lt}(p,\gamma_0) \le -c$, for $\gamma \in \Gamma$ and $p \in \mathcal{P}_{lt}$. Therefore, by Taylor expansion of $r_{ls}(p,\gamma)$ around p^*_{ls}, we obtain

$$r_{ls}(p_{ls},\gamma_0) = r_{ls}(p^*_{ls}(\gamma_0),\gamma_0) + r'_{ls}(p^*_{ls}(\gamma_0),\gamma_0)(p_{ls}-p^*_{ls}(\gamma_0)) + \frac{1}{2} r''_{ls}(\tilde{p},\gamma_0)(p_{ls}-p^*_{ls}(\gamma_0))^2\,,$$

for some $\tilde{p}$ between p_{ls} and p^*_{ls}. By optimality of p^*_{ls} we have $r'_{ls}(p^*_{ls}(\gamma_0),\gamma_0) = 0$. In addition, since $\tilde{p} \in \mathcal{P}_{ls}$, we have $r''_{ls}(\tilde{p},\gamma_0) < -c$, which implies that

$$(p_{ls}-p^*_{ls}(\gamma_0))^2 \le \frac{2}{c}\Big(r_{ls}(p^*_{ls}(\gamma_0),\gamma_0) - r_{ls}(p_{ls},\gamma_0)\Big)\,.$$

Using the above bound in (37), we arrive at

$$\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) \le C(\gamma-\gamma_0)^2 \sum_{s=1}^{t}\sum_{l=1}^{L} n_{ls}\mathbb{E}_{\gamma_0}[r_{ls}(p^*_{ls}(\gamma_0),\gamma_0) - r_{ls}(p_{ls},\gamma_0)] \le C(\gamma-\gamma_0)^2\, \mathrm{Reg}_t\,, \tag{38}$$

which completes the proof (28).

We next proceed with our proof for bound (29). Recall the optimality condition

$$v_l\gamma p^*_{lt}(\gamma) = \frac{\Phi(v_l\gamma(x_{lt}-p^*_{lt}(\gamma)))}{\phi(v_l\gamma(x_{lt}-p^*_{lt}(\gamma)))}\,.$$

Differentiating with respect to γ on both sides we get,

$$v_l p^*_{lt}(\gamma) + v_l\gamma\frac{d}{d\gamma}p^*_{lt}(\gamma) = v_l\left(x_{lt} - p^*_{lt}(\gamma) - \gamma\frac{d}{d\gamma}p^*_{lt}(\gamma)\right)\kappa(\gamma)\,,$$

where

$$\kappa(\gamma) = \frac{\phi^2(v_l\gamma(x_{lt} - p^*_{lt}(\gamma))) - \Phi(v_l\gamma(x_{lt} - p^*_{lt}(\gamma)))\phi'(v_l\gamma(x_{lt} - p^*_{lt}(\gamma)))}{\phi^2(v_l\gamma(x_{lt} - p^*_{lt}(\gamma)))}\,.$$

By rearranging the terms we have

$$\frac{d}{d\gamma}p^*_{lt}(\gamma) = \frac{1}{\gamma}\left(-p^*_{lt}(\gamma) + \frac{k(\gamma)}{1+k(\gamma)}\right).$$

Since we are working on finite sets, we can restrict the problem class Γ, such that $|\frac{d}{d\gamma}p(\gamma)| > C$, for some constant C and all $\gamma \in \Gamma$. Therefore, by an application of the Mean Value Theorem, we have

$$|p^*_{lt}(\gamma) - p^*_{lt}(\gamma_0)| \geq C|\gamma - \gamma_0|\,.$$

Let $\gamma_1 := \gamma_0 + 1/(4T^{1/4})$. Using the above bound, the optimal prices for γ_0 and γ_1 are apart by at least $C/(4T^{1/4})$.

Consider two disjoint sets D_1 and D_0 of prices, as follows:

$$D_{\gamma_0} := \left\{p : |p - p^*_{lt}(\gamma_0)| \leq \frac{C}{10T^{1/4}}\right\}\,, \qquad D_{\gamma_1} := \left\{p : |p - p^*_{lt}(\gamma_1)| \leq \frac{C}{10T^{1/4}}\right\}.$$

Note that D_{γ_0} and D_{γ_1} are disjoint since $|p^*_{lt}(\gamma_1) - p^*_{lt}(\gamma_0)| \geq C/(4T^{1/4})$.

For $\gamma \in \{\gamma_0, \gamma_1\}$, if the posted price p_{lt} is not in the set D_γ, then the instantaneous regret is at least

$$r_{lt}(p^*_{lt}(\gamma), \gamma) - r_{lt}(p_{lt}, \gamma) \geq \frac{c}{2}(p^*_{lt}(\gamma) - p_{lt}) \geq \left(\frac{cC}{20}\right)^2 \frac{1}{\sqrt{T}}\,.$$

Hence following a similar proof strategy as in (Broder & Rusmevichientong 2012, Lemma 3.4), we have

$$\mathrm{Reg}^{\pi,\gamma_0}_T + \mathrm{Reg}^{\pi,\gamma_1}_T \geq \left(\frac{cC}{20}\right)^2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T}\sum_{l=1}^{L} n_{lt}\left(\mathbb{P}_{\gamma_0}(p_{lt} \notin D_{\gamma_0}) + \mathbb{P}_{\gamma_1}(p_{lt} \notin D_{\gamma_1})\right)\,.$$

Note that p_{lt} is measurable with respect to measure $f^{\pi,\gamma}_{t-1}$ under the model γ. Therefore, by using a standard result on the minimum error in a simple hypothesis test (Tsybakov 2004, Theorem

2.2), we have

$$
\begin{aligned}
\mathrm{Reg}_T^{\pi,\gamma_0} + \mathrm{Reg}_T^{\pi,\gamma_1} &\geq \frac{C_1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{l=1}^{L} n_{lt} e^{-\mathsf{KL}\left(f_{t-1}^{\pi,\gamma_0}; f_{t-1}^{\pi,\gamma_1}\right)} \\
&\geq \frac{C_1}{\sqrt{T}} N_T e^{-\mathsf{KL}\left(f_T^{\pi,\gamma_0}; f_T^{\pi,\gamma_1}\right)}, \qquad (39)
\end{aligned}
$$

where in the second step we used the fact that $\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma_1}\right)$ is non-decreasing in t and $N_T := \sum_{t=1}^{T}\sum_{l=1}^{L} n_{lt}$. Previously we established the lower bound (38), which reads as

$$
\mathrm{Reg}_T^{\pi,\gamma_0} \geq \frac{C_2}{(\gamma_0 - \gamma)^2} \mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma}\right) .
$$

Putting, $\gamma = \gamma_1 = \gamma_0 + 1/4T^{1/4}$ we get $\mathrm{Reg}_T^{\pi,\gamma_0} \geq C_2\sqrt{T}\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma_1}\right)$. Combining this bound with (39), we get

$$
\begin{aligned}
\max_{\gamma\in\{\gamma_0,\gamma_1\}} \mathrm{Reg}_T^{\pi,\gamma} &\geq \frac{1}{2}\left(\mathrm{Reg}_T^{\pi,\gamma_0} + \mathrm{Reg}_T^{\pi,\gamma_1}\right) \\
&\geq C\sqrt{T}\left(\mathsf{KL}\left(f_t^{\pi,\gamma_0}; f_t^{\pi,\gamma_1}\right) + \frac{N_T}{T} e^{-\mathsf{KL}\left(f_T^{\pi,\gamma_0}; f_T^{\pi,\gamma_1}\right)}\right) \\
&\geq C\sqrt{T}\left(1 + \log(N_T/T)\right) ,
\end{aligned}
$$

where in the last step we used the inequality $ae^{-b} + b \geq 1 + \log(a)$.

C Proofs of all other results and intermediate steps

C.1 Proof of Proposition 2.1

Recall that $V_{lt}^2 = \|(\boldsymbol{I} - \rho_t \boldsymbol{W})^{-1}\boldsymbol{e}_l\|^2\tau^2 + \sigma^2$. Since $W \succeq \boldsymbol{0}$, and $\rho_t \geq 0$, we have $\boldsymbol{I} - \rho_t\boldsymbol{W} \preceq \boldsymbol{I}$. In addition, by Assumption 2.2, we have $\boldsymbol{I} - \rho_t\boldsymbol{W} \succeq \varepsilon\boldsymbol{I}$. Therefore, since $\|\boldsymbol{e}_l\| = 1$,

$$
1 \leq \|(I - \rho_t\boldsymbol{W})^{-1}\boldsymbol{e}_l\| \leq \frac{1}{\varepsilon} ,
$$

from which we obtain the result.

Next, to prove the upper bound on the optimal prices, we recall that

$$0 \le \boldsymbol{x}_{lt}'\boldsymbol{m}_{lt} \le \|\boldsymbol{x}_{lt}\|\|\boldsymbol{m}_{lt}\| \le \frac{\|\boldsymbol{\mu}_t\|}{V_{lt}} \le \frac{C_\mu}{c_V}\,.$$

Invoking relation (14), and noting that β_t and so b_t are negative, we arrive at

$$p_{lt}^* = \frac{1}{-b_{lt}}\left(\varphi^{-1}(-\boldsymbol{x}_{lt}'\boldsymbol{m}_{lt}) + \boldsymbol{x}_{lt}'\boldsymbol{m}_{lt}\right) \le c_\beta^{-1} C_V (C_\mu c_V^{-1} - 0.5\phi(0))\,,$$

where we used that φ^{-1} is increasing, $\boldsymbol{x}_{lt} \ge 0$, and $\varphi^{-1}(0) = -0.5/\phi(0)$.

C.2 Proof of Lemma 3.2

By definition, $V_{lt}^2 = \|(\boldsymbol{I} - \rho_t \boldsymbol{W})^{-1}\boldsymbol{e}_l\|^2\tau^2 + \sigma^2$. Hence, if ω_* is the smallest eigenvalue of $\boldsymbol{W}$, then $V_{lt}^2 \ge \tau^2/(1-\rho_t\omega_*)^2$.

We want to bound the $|b_{l,t+1} - b_{l,t}|$ and $\|\boldsymbol{m}_{l,t+1} - \boldsymbol{m}_{lt}\|_2$.

$$\begin{aligned}
\|\boldsymbol{m}_{l,t+1} - \boldsymbol{m}_{lt}\|_2 &= \left\|\frac{\boldsymbol{\mu}_{t+1}}{V_{l,t+1}} - \frac{\boldsymbol{\mu}_t}{V_{lt}}\right\|_2 \\
&\le \left\|\frac{\boldsymbol{\mu}_{t+1} - \boldsymbol{\mu}_t}{V_{l,t+1}}\right\|_2 + \boldsymbol{\mu}_t\left\{\frac{1}{V_{l,t+1}} - \frac{1}{V_{lt}}\right\} \\
&\le \frac{\delta_{t\mu}}{\tau/(1-\rho_t\omega_*)} + C_\mu\left\{\frac{1}{V_{l,t+1}} - \frac{1}{V_{lt}}\right\}.
\end{aligned}$$

Further, the second term can be simplified as

$$\begin{aligned}
\left\{\frac{1}{V_{l,t+1}} - \frac{1}{V_{lt}}\right\} = \frac{V_{lt} - V_{l,t+1}}{V_{l,t+1}V_{lt}} &= \frac{V_{lt}^2 - V_{l,t+1}^2}{V_{l,t+1}V_{lt}(V_{lt} + V_{l,t+1})} \\
&\le \frac{1}{2c_V^3}(V_{lt}^2 - V_{l,t+1}^2) \\
&\le \frac{\tau^2}{2c_V^3}(\|(\boldsymbol{I} - \rho_t\boldsymbol{W})^{-1}\boldsymbol{e}_l\|^2 - \|(\boldsymbol{I} - \rho_{t+1}\boldsymbol{W})^{-1}\boldsymbol{e}_l\|^2) \le C\delta_{t\rho}\,.
\end{aligned}$$

The same analysis can be done for $|b_{l,t+1} - b_{lt}|$ as well.

C.3 Proof of Corollary 3.3

The corollary follows directly by applying the results from Lemma 3.2 in Theorem 3.1. Since $\rho_t = \rho$ for all t, hence $\delta_{t\rho} = 0$.

Since $\eta_t \propto 1/\sqrt{t}$, we get

- $\mathcal{R}_1 = LC_1 C\tau^{-1}(1-\rho\omega_*) \sum_{t=1}^{T} \sqrt{t}\delta_{t\beta}$,
- $\mathcal{R}_2 = LC_1 C\tau^{-1}(1-\rho\omega_*) \sum_{t=1}^{T} \sqrt{t}\delta_{t\mu}$,
- $\mathcal{R}_3 = C_3 C \sum_{t=1}^{T} n_t^2/\sqrt{t} = \mathcal{O}(\sqrt{T})$,
- $\mathcal{R}_4 = C_4 CL(C_b + C_m)\sqrt{T+1} = \mathcal{O}(\sqrt{T})\,.$

Changing the constants appropriately gives us the corollary.

C.4 Proof of Lemma 3.5

Consider the particular set-up when $\boldsymbol{\mu}_t = 0$, $\beta_t = \beta$, $\rho_t = \rho$ and $\sigma = 1$ in (2). Further, assume $n_{lt} = n$ for all l, t and $n \to \infty$. The proof can easily be extended to the generic set-up. Under these parametric assumptions, first note that, $\text{Rev}(\boldsymbol{\lambda}, l, t, p_{lt}) = np_{lt}\Phi(\alpha_{lt} + \beta p_{lt})$. The optimal pricing strategy p_{lt}^* maximizes $\text{Rev}(\boldsymbol{\lambda}, l, t, p_{lt})$ over p_{lt} for any fixed $\boldsymbol{\lambda}$.

Now, consider an arbitrary pricing policy $\boldsymbol{p}$ based on the unpenalized likelihood $\text{PL}(\boldsymbol{\lambda}, 0)$. Such a policy will be dominated by its oracle counter-part $\boldsymbol{p}^{\text{or}}$ which already knows the price coefficient β and also, knows the latent utility U_{lt}. Note that, the revenue of any pricing policy $\boldsymbol{p}$ based on the unpenalized likelihood is always dominated by the revenue of this oracle strategy, i.e., $\text{Rev}(\boldsymbol{\lambda}, \boldsymbol{p}) \leq \text{Rev}(\boldsymbol{\lambda}, \boldsymbol{p}^{\text{or}})$. Subsequently, the oracle strategy $\boldsymbol{p}^{\text{or}}$ will have a lower regret. Next, we concentrate on the regret of $\boldsymbol{p}^{\text{or}}$.

For this calculation note that based on model (2), the only unknown parameters for the oracle strategy $\boldsymbol{p}^{\text{or}}$ are the α_{lt}s. Under this framework consider α_{lt}s being best estimated by $\hat{\alpha}_{lt}^{\text{or}}$.

Now, note that as we do not have any structural assumption between $\boldsymbol{\alpha}_t$ and $\boldsymbol{\alpha}_{t+1}$ over $t = 1, \ldots, T$, for any t, $\hat{\alpha}_{lt}^{\text{or}}$ will be estimated based on $\{U_{ltk} : l = 1, \ldots, L; k = 1, \ldots, n\}$. As the prices p_{lt} are known (based on the filtration $\mathcal{F}_{t-1}$ which contains all information upto time $t-1$)

this further reduces to estimating the the L means $\boldsymbol{\alpha}_t$ from uncorrelated L dimensional Gaussian location model where we observe $n^{-1}\sum_{k=1}^{n} U_{ltk} - \beta p_{lt}$ for $l = 1, \ldots, L$. From the Cramer-Rao lower bound for Gaussian family, it follows that for all $l = 1, \ldots, L$, we will have the following error bound on any estimate $\hat{\alpha}_{lt}$:

$$\mathbb{E}_{\boldsymbol{\lambda}}(\hat{\alpha}_{lt} - \alpha_{lt})^2 \geq n^{-1}.$$

As such consider the α_{lt}s under the oracle framework to be estimated by the MLE. Let $\hat{\delta}_{lt} = \hat{\alpha}^{\text{or}}_{lt} - \alpha_{lt}$. Then, noting that the MLE is asymptotically rotation invariant in this case, we have for any $\boldsymbol{\lambda}$:

$$\mathbb{E}_{\boldsymbol{\lambda}}\, \hat{\boldsymbol{\delta}}_t \hat{\boldsymbol{\delta}}_t^T = n^{-1} I_L \;\text{ and } \mathbb{E}_{\boldsymbol{\lambda}}\, \hat{\boldsymbol{\delta}}_t \to \mathbf{0} \text{ as } n \to \infty. \tag{40}$$

Now, note that for the oracle strategy,

$$\text{Rev}(\boldsymbol{\lambda}, l, t, p^{\text{or}}_{lt}) = \max_{p \geq 0} np\, \Phi(\hat{\alpha}^{\text{or}}_{lt} + \beta p) = \max_{p \geq 0} np\, \Phi(\hat{\delta}_{lt} + \alpha_{lt} + \beta p)\ .$$

Let $f(l, t, p) = np\, \Phi(\hat{\delta}_{lt} + \alpha_{lt} + \beta p)$. Consider Taylor-Series expansion:

$$f(l, t, p) = np\, \Phi(\alpha_{lt} + \beta p) + n\, \hat{\delta}_{lt}\, p\, \phi(\alpha_{lt} + \beta p) + 2^{-1} np\, \hat{\delta}^2_{lt} \phi'(\alpha_{lt} + \beta p) + r(l, t, p),$$

where $r(l, t, p)$ contains third and higher order terms. Now, we have $L^{-1}\sum_l f(l, t, p_{lt})$ converges in probability to

$$\frac{1}{L}\sum_l np_{lt}\, \Phi(\alpha_{lt} + \beta p_{lt}) + \frac{1}{L}\sum_l n\, p_{lt} \phi(\alpha_{lt} + \beta p_{lt}) \mathbb{E}_{\boldsymbol{\lambda}} \hat{\delta}_{lt} + \frac{1}{2L}\sum_l np_{lt} \phi'(\alpha_{lt} + \beta p_{lt}) \mathbb{E}_{\boldsymbol{\lambda}} \hat{\delta}^2_{lt},$$

as $L^{-1}\sum_l r(l, t, p_{lt}) \to 0$ in probability as $n\, \mathbb{E}_{\boldsymbol{\lambda}} \hat{\delta}^{2+m}_{lt} = O(n^{-m/2})$, for $m \geq 1$. Using (40), the second term in the above expression vanishes and the third term gets further simplified, resulting in the following asymptotic result:

$$\frac{1}{L}\sum_{l} f(l,t,p_{lt}) = n\left[\frac{1}{L}\sum_{l} p_{lt}\,\Phi(\alpha_{lt}+\beta p_{lt})\right] + \frac{1}{2L}\sum_{l} p_{lt}\phi'(\alpha_{lt}+\beta p_{lt}) + o(1)\ .$$

Thus, the regret of $\boldsymbol{p}^{\mathrm{or}}$ at time t is given by

$$L^{-1}\sum_{l=1}^{L}\mathcal{R}_{lt}(\boldsymbol{\lambda},\boldsymbol{p}^{\mathrm{or}}) \geq (\mathcal{A}-\mathcal{B})/L + o(1), \tag{41}$$

where,

$$\begin{aligned}
\mathcal{A} &= \max_{p_{lt}:l=1,\ldots,L}\left[\sum_{l} np_{lt}\,\Phi(\alpha_{lt}+\beta p_{lt})\right],\ \text{and}\\
\mathcal{B} &= \max_{p_{lt}:l=1,\ldots,L}\left[\sum_{l} np_{lt}\,\Phi(\alpha_{lt}+\beta p_{lt}) - 2^{-1}\sum_{l} p_{lt}(\alpha_{lt}+\beta p_{lt})\phi(\alpha_{lt}+\beta p_{lt})\right].
\end{aligned}$$

Note that, the expression in $\mathcal{B}$ is simplified using $\phi'(u) = -u\phi(u)$. Now recall that β, being the price sensitivity, is negative. Based on model (2), for the utilities to be positive we have the following assumption of the price: $\alpha_{lt}+\beta p_{lt} > 0$ for all l and t. Let the prices be selected such that $\inf_l \alpha_{lt}+\beta p_{lt} > \epsilon_0$ for some prefixed small $\epsilon_0 > 0$. By Proposition 2.1, the optimal prices are bounded and so are $\sup_l \alpha_{lt}+\beta p_{lt} < M_0$. Then,

$$\mathcal{A}-\mathcal{B} \geq 2^{-1}\epsilon\sum_{l} p^{*}_{lt},$$

where p^{*}_{lt} is the optimal price based on criterion $\mathcal{B}$, and $\epsilon = \min_{\epsilon_0<|u|<M'} u\phi(u)$. Thus, the cumulative regret of $\boldsymbol{p}^{\mathrm{or}}$ over time is given by

$$\mathcal{R}(\boldsymbol{\lambda},\boldsymbol{p}^{\mathrm{or}}) = \sum_{t=1}^{T}\sum_{l=1}^{L}\mathcal{R}_{lt}(\boldsymbol{\lambda},\boldsymbol{p}^{\mathrm{or}}) = \Omega(LT).$$

Thus, we have,

$$\mathcal{B}(\boldsymbol{\theta},\boldsymbol{p}_U) = \Omega(LT).$$

Now, consider the regret from the proposed strategy. Based on (24), we have

$$\mathcal{B}(\boldsymbol{\theta}, \boldsymbol{p}) \leq C_6 \tau^{-1}(1-\rho_* \omega_*) \sum_{t=1}^{T} \sqrt{t}(\delta_{t\beta} + \delta_{t\mu}) + C_7 \sum_{t=1}^{T} \sqrt{t}\delta_{t\rho} + \mathcal{O}(\sqrt{T}) = \mathcal{O}(\sqrt{T}),$$

where, the second asymptotic result follows as $\sum_{t=1}^{T} \sqrt{t}\delta_{t\beta}$, $\sum_{t=1}^{T} \sqrt{t}\delta_{t\mu}$ and $\sum_{t=1}^{T} \sqrt{t}\delta_{t\rho}$ are all bounded above by $\mathcal{O}(\sqrt{T})$. Comparing the above two displays the result follows.

C.5 Proof of Proposition 4.1

Consider the revenue function given by (25), $\mathrm{Rev}_{lt}(p) = n_{lt} p \Phi(b_{lt} p + \boldsymbol{x}_{lt}' \boldsymbol{m}_{lt})$. By definition, p_{lt}^* is the maximizer of $\mathrm{Rev}_{lt}(p)$ and hence $\mathrm{Rev}_{lt}'(p_{lt}^*) = 0$. Using Taylor series expansion around p_{lt}^*, we get

$$\mathcal{R}_{lt} = \mathrm{Rev}_{lt}(p_{lt}^*) - \mathrm{Rev}_{lt}(p_{lt}) = \frac{1}{2}\mathrm{Rev}_{lt}''(p)(p_{lt} - p_{lt}^*)^2,$$

for some p between p_{lt} and p_{lt}^*. The second derivative can be bounded as

$$\frac{1}{2}\mathrm{Rev}_{lt}''(p) = n_{lt} \frac{2 b_{lt} \phi(b_{lt} p + \boldsymbol{x}_{lt}' \boldsymbol{m}_{lt}) + p b_{lt}^2 \phi'(b_{lt} p + \boldsymbol{x}_{lt}' \boldsymbol{m}_{lt})}{2} \leq \left(C_b \phi(0) + M C_b^2 \frac{\phi(0)}{\sqrt{2}} \right) n_{lt}.$$

Setting $C_9 = C_b \phi(0) + M C_b^2 \frac{\phi(0)}{\sqrt{2}}$ completes the proof.

C.6 Proof of Lemma 4.2

We have the true parameters $b_{lt}, \boldsymbol{m}_{lt}$ and the output $\hat{b}_{lt}, \hat{\boldsymbol{m}}_{lt}$ from our PSGD pricing policy. Also, p_{lt} and p_{lt}^* are the price based on our policy and the optimal price based on the true parameters, respectively.

As discussed in section 2, we can write the prices in terms of the utility model parameters using the function $g(\cdot,\cdot)$, as follows:

$$\begin{aligned}
p^*_{lt} &:= g(b_{lt}, \boldsymbol{m}_{lt}) = -\frac{\varphi^{-1}(-\boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) + \boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}}{b_{lt}}\ , \\
p_{lt} &:= g(\hat{b}_{lt}, \hat{\boldsymbol{m}}_{lt}) = -\frac{\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}}{\hat{b}_{lt}}\ .
\end{aligned}$$

Now, note that,

$$(p_{lt} - p^*_{lt})^2 = \Big\{(\varphi^{-1}(-\boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) + \boldsymbol{x}'_{lt}\boldsymbol{m}_{lt})\, {b_{lt}}^{-1} - (\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt})\, {\hat{b}_{lt}}^{-1}\Big\}^2 = \{A+B\}^2,$$

where, the right side above is decomposed as

$$\begin{aligned}
A &= (\varphi^{-1}(-\boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) + \boldsymbol{x}'_{lt}\boldsymbol{m}_{lt})\, {b_{lt}}^{-1} - (\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt})\, {b_{lt}}^{-1},\ \text{and,} \\
B &= \big(\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}\big)\ (1/b_{lt} - 1/\hat{b}_{lt}).
\end{aligned}$$

Using the naive bound $\{A+B\}^2 \le 2(A^2+B^2)$ first and then the bound $|b_{lt}| = |\beta_t|/V_{lt} \ge c_\beta/C_V$ and the policy rule $p_{lt} = g(\hat{b}_{lt}, \hat{\boldsymbol{m}}_{lt})$, it follows that $(p_{lt} - p^*_{lt})^2$ is bounded above by

$$2C_V^2 c_\beta^{-2}\left\{(\varphi^{-1}(-\boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) + \boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) - (\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt})\right\}^2 + 2p_{lt}^2 {b_{lt}}^{-2}(b_{lt} - \hat{b}_{lt})^2\,,$$

Since, $\varphi^{-1}(-v) + v$ is 1-Lipschitz, we have

$$\big((\varphi^{-1}(-\boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) + \boldsymbol{x}'_{lt}\boldsymbol{m}_{lt}) - (\varphi^{-1}(-\boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt}) + \boldsymbol{x}'_{lt}\hat{\boldsymbol{m}}_{lt})\big)^2 \le \langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt}\rangle^2\,.$$

Therefore, we have

$$(p_{lt} - p^*_{lt})^2 \le 2C_V^2 c_\beta^{-2}\langle \boldsymbol{x}_{lt}, \boldsymbol{m}_{lt} - \hat{\boldsymbol{m}}_{lt}\rangle^2 + 2C_V^2 c_\beta^{-2} p_{lt}^2 (b_{lt} - \hat{b}_{lt})^2\,.$$

Setting $C_{10} = 2C_V^2 c_\beta^{-2}$ proves the lemma.

D Further Numerical Experiments

Set-up 8. Next, we want to study the regret behaviour under a different network on the segments. Unlike Set-ups 3 and 4, we create a network on the US states based only on the demographic variables. Based on the demographic variables, we create the similarity matrix $\boldsymbol{W}$ using an RBF kernel of width two and threshold the edges at 0.05. Figure 9 shows the network.

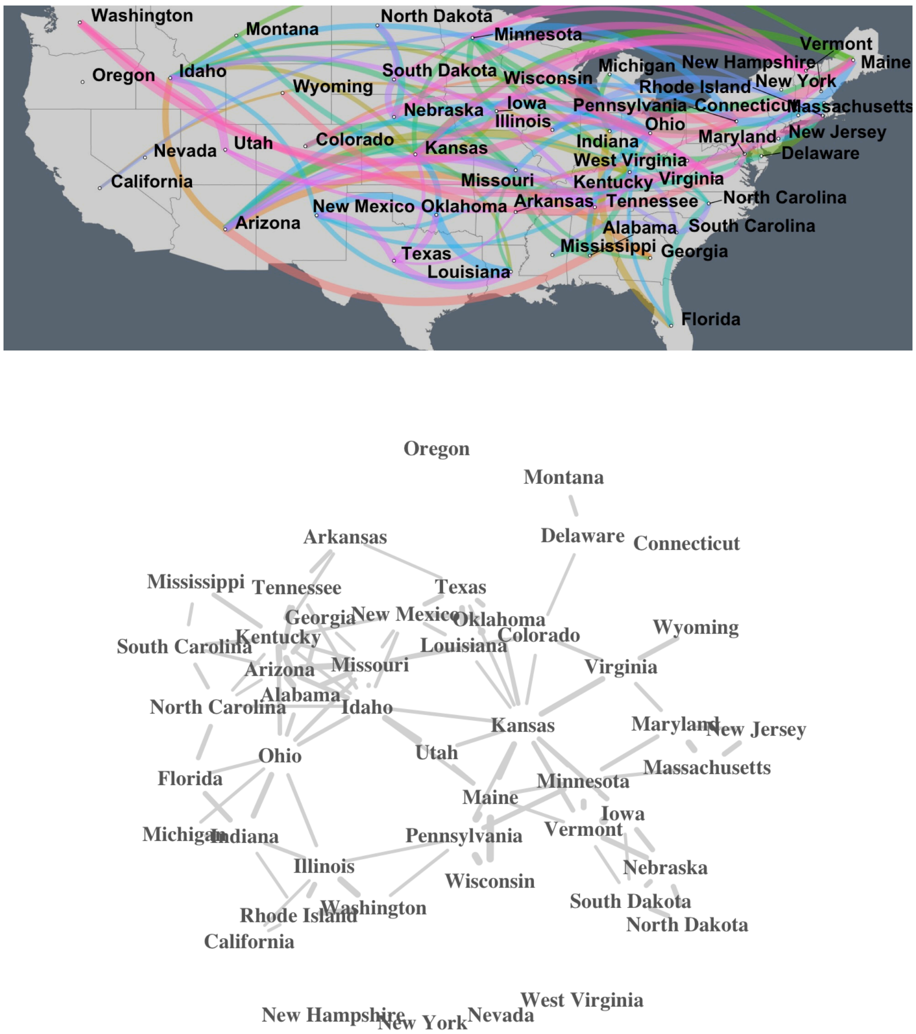


Figure 9: A network on 48 US states (barring Hawaii and Alaska) based on demographic variables. It is used in set-up 5.

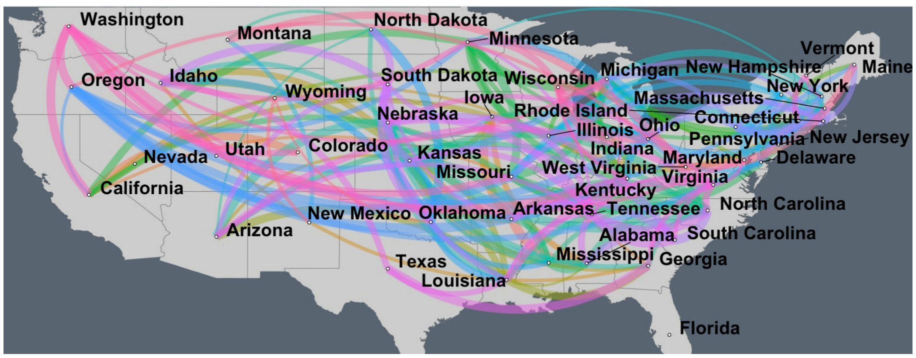
We study the regret of the imbalanced design in this network regime. Similar to the set-up 4, we compare our policy in the imbalanced setting against (a) unshrunken policy in the balanced

Table 3: Performance of PSGD in Set-up 5 relative to (a) unshrunken policy in the balanced designs and (b) unshrunken policy in the imbalanced designs (negative implies worse performance)

n	T	0.7		0.8		0.9	
		Balanced	**Imbalanced**	**Balanced**	**Imbalanced**	**Balanced**	**Imbalanced**
1000	**100**	-9.1%	-12.4%	-8.2%	-21.7%	-6.1%	-34.0%
	500	-13.5%	-16.9%	-18.0%	-35.5%	-20.8%	-59.1%
	1000	-1.0%	-4.1%	-3.2%	-20.5%	-4.4%	-42.7%
	5000	54.8%	53.4%	54.3%	50.0%	55.7%	49.8%
2500	**100**	-5.4%	-8.6%	-2.1%	-21.1%	3.2%	-32.5%
	500	-13.8%	-17.2%	-10.2%	-27.4%	-10.8%	-53.6%
	1000	-3.9%	-7.0%	0.5%	-13.8%	0.9%	-40.0%
	5000	45.1%	43.4%	48.7%	49.9%	51.5%	48.8%
5000	**100**	-8.0%	-11.2%	-7.3%	-19.3%	-2.3%	-31.3%
	500	-15.6%	-19.0%	-15.9%	-27.9%	-13.2%	-45.5%
	1000	-3.8%	-6.9%	-3.3%	-13.9%	0.7%	-30.3%
	5000	46.8%	45.2%	48.3%	49.7%	52.2%	52.3%
10000	**100**	-7.9%	-11.2%	-4.3%	-22.1%	0.6%	-30.0%
	500	-12.4%	-15.8%	-12.0%	-30.1%	-10.4%	-47.9%
	1000	-0.4%	-3.4%	1.3%	-15.7%	3.3%	-34.5%
	5000	48.5%	47.0%	51.0%	50.6%	54.0%	51.5%
20000	**100**	-6.9%	-10.1%	-2.4%	-21.0%	1.8%	-31.3%
	500	-13.9%	-17.3%	-10.4%	-29.6%	-10.8%	-49.6%
	1000	-2.1%	-5.2%	1.9%	-16.0%	2.2%	-36.2%
	5000	46.2%	44.6%	49.9%	48.8%	52.5%	50.6%

setup and (b) unshrunken policy in the imbalanced setting. The results are presented in Table 3. At smaller T's our policy performs worse than an unshrunken policy, but as T grows larger, our policy performs significantly better (more than 50%) compared to the unshrunken policy, both with balanced and imbalanced design.

Set-up 9. We use a network that is based on the similarity across economic variables only. We create the network of the US states based only on the economic variables using an RBF kernel of width two and thresholding the edges at 0.05. Figure 10 shows the network.



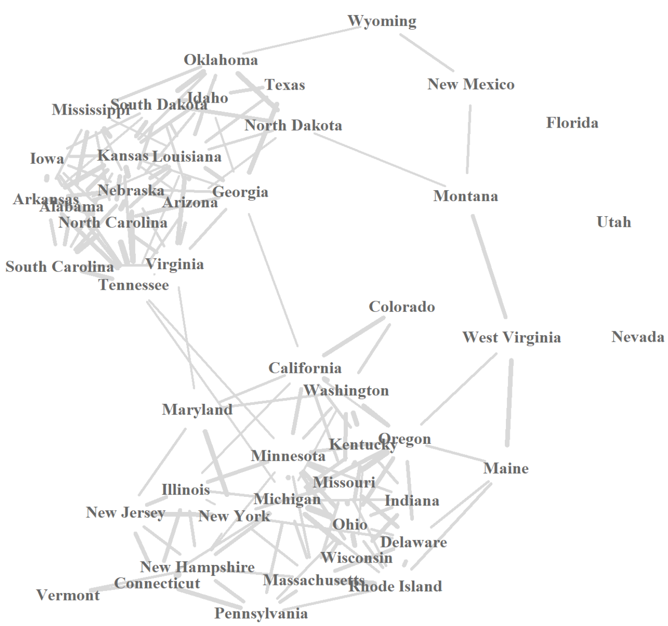


Figure 10: A network on 48 US states (barring Hawaii and Alaska) based on economic variables. It is used in set-up 6.

We study the performance under this network regime in the imbalanced setting with the two scenarios as above. The results are reported in Table 4. We see that overall, in all imbalanced settings our policy performs much better than the unshrunken policy in the imbalanced as well as the balanced setting.

Table 4: Performance of prescribed method in Set-up 6 compared to (a) unshrunken policy in the balanced setup and (b) unshrunken policy in the imbalanced setup (negative implies worse performance)

n	**T**	**0.7**		**0.8**		**0.9**	
		Balanced	**Imbalanced**	**Balanced**	**Imbalanced**	**Balanced**	**Imbalanced**
1000	**100**	-4.2%	-2.1%	-7.3%	-2.3%	-4.3%	-3.8%
	500	4.9%	6.8%	2.0%	6.6%	3.6%	4.3%
	1000	10.7%	12.5%	8.8%	12.6%	10.3%	10.3%
	5000	29.3%	30.7%	27.7%	30.8%	29.1%	29.2%
2500	**100**	-6.2%	-4.0%	-6.2%	-3.0%	-9.4%	-5.9%
	500	4.8%	6.7%	3.4%	6.9%	-1.1%	3.0%
	1000	11.7%	13.5%	10.1%	12.9%	6.4%	9.4%
	5000	30.8%	32.2%	28.9%	31.2%	25.9%	28.5%
5000	**100**	-7.4%	-5.2%	-6.1%	-2.5%	-9.8%	-5.6%
	500	3.7%	5.6%	3.4%	6.9%	-1.0%	3.0%
	1000	10.2%	12.0%	10.2%	13.0%	6.2%	9.5%
	5000	29.9%	31.3%	29.1%	31.4%	25.9%	28.7%
10000	**100**	-6.5%	-4.4%	-2.4%	-2.8%	-4.7%	-6.1%
	500	4.5%	6.4%	7.0%	6.7%	3.7%	2.9%
	1000	11.0%	12.8%	13.4%	12.8%	10.8%	9.4%
	5000	30.0%	31.4%	31.4%	31.0%	29.8%	28.9%
20000	**100**	-5.8%	-3.7%	-3.5%	-2.8%	-6.8%	-5.7%
	500	4.6%	6.5%	6.1%	6.9%	1.6%	3.0%
	1000	11.0%	12.8%	12.7%	13.1%	8.7%	9.3%
	5000	30.2%	31.6%	31.0%	31.4%	28.0%	28.6%