A Theoretical Details

We will include omitted technical details in the main context. We first summarize all the required additional assumptions in Section A.1. Then, we provide omitted proofs for Section 3 in Section A.2 and omitted proofs for Section 4 in Section A.3.

A.1 Assumptions

In this subsection, we summarize all the additional assumptions we will use to build various theoretical results in this paper. Before we state the assumptions, we need some further notations.

Let us denote $\Psi(\theta)$ as the collection of minimizers of $\int p(z,\theta)\|\nabla_\theta \log p(z,\theta) - s(z,\theta;\psi)\|^2 dz$ for the given θ. We further denote the empirical estimation function for the two terms, i.e.,

$$\int p(z,\theta)\Big(\|s(z,\theta;\psi)\|^2 - 2\nabla_\theta \log p(z,\theta)^\top s(z,\theta;\psi)\Big)dz,$$

as $\hat{J}_{n,k}(\psi;\theta)$ following the methods in Section 3.2 for any θ in the trajectory $\{\hat{\theta}_t\}_{t=1}^T$, where n and k are the number of samples we get at each iteration for $\hat{\theta}_t$ and perturbed policies.

Meaning of each assumption. Assumption A.1 is used to establish the asymptotic normality of our estimator under performativity. Additionally, we use the fact that the population loss Hessians $H_{\tilde{\theta}}(\theta) = \nabla_\theta^2 \mathbb{E}_{z\sim\mathcal{D}(\tilde{\theta})}\ell(z;\theta)$ are positive definite, which is guaranteed by the strong convexity of the loss function (Assumption 3.1). Assumption A.2 and A.3 are used in the analysis of score matching. Assumption A.2, based on the differentiation lemma [16], ensures the interchangeability of integration and differentiation. Assumption A.3 guarantees the consistency of the score matching estimator. Assumption A.4 and A.5 parallel to Assumption 3.1(a) and A.1, and are used to establish the consistency and asymptotic normality of our PPI estimator under performativity. Conditions such as local Lipschitz continuity and positive definiteness are standard for establishing asymptotic normality. Similar assumptions are also imposed in [1].

Assumption A.1 (Positive Definiteness & Regularity Conditions for the Estimator)**.** We assume the following.

(a). The loss function satisfies the the gradient covariance matrices are positive definite:

$$V_{\tilde{\theta}}(\theta) = \mathrm{Cov}_{z\sim\mathcal{D}(\tilde{\theta})}\big(\nabla_\theta \ell(z;\theta)\big) \succ 0,$$

for any $\tilde{\theta},\theta \in \Theta$.

(b). For any sample size n, assume the M-estimator $\hat{\theta}_t$ has a density function with respect to the Lebesgue measure, and its characteristic function is absolutely integrable.

Assumption A.2 (Regularity Condition for M)**.** Assume that for $\forall i$:

(a). The function $z \mapsto p(z,\theta)\dfrac{\partial \log M(z,\theta;\psi)}{\partial \theta^{(i)}}$ is Lebesgue integrable.

(b). For almost every $z \in \mathcal{Z}$ (with respect to Lebesgue measure), the partial derivative $\dfrac{\partial}{\partial \theta^{(i)}}\left[p(z,\theta)\dfrac{\partial \log M(z,\theta;\psi)}{\partial \theta^{(i)}}\right]$ exists.

(c). There exists a Lebesgue-integrable function $H(z)$ such that for almost every $z \in \mathcal{Z}$,

$$\left|\frac{\partial}{\partial \theta^{(i)}}\left[p(z,\theta)\frac{\partial \log M(z,\theta;\psi)}{\partial \theta^{(i)}}\right]\right| \le H(z).$$

Assumption A.3 (Consistency of Optimizer)**.** We let k grows along with n such that $n \to \infty$ leads to $k \to \infty$. We assume that the class $M(z,\theta;\psi)$ is rich enough that for all $\theta \in \Theta$, there exists $\psi^*(\theta)$ such that $M(z,\theta;\psi^*(\theta)) = p(z,\theta)$. Moreover, for the underlying trajectory $\{\theta_t\}_{t=1}^T$,

$$\lim_{n\to\infty} \mathrm{argmin}_\psi\, \hat{J}_{n,k}(\psi;\hat{\theta}_t) \subseteq \Psi(\theta_t).$$

Assumption A.4 (Local Lipschitzness with f)**.** Loss function $\ell(x,f(x);\theta)$ is locally Lipschitz: for each $\theta \in \Theta$, there exist a neighborhood $\Upsilon(\theta)$ of θ such that $\ell(x,f(x);\tilde{\theta})$ is $L^f(x)$ Lipschitz w.r.t $\tilde{\theta}$ for all $\tilde{\theta} \in \Upsilon(\theta)$ and $\mathbb{E}_{x\sim\mathcal{D}_\mathcal{X}(\theta)}L^f(x) < \infty$.

Assumption A.5 (Positive Definiteness with f & Regularity Conditions for the PPI Estimator)**.** We assume the following.

(a). Assume the loss function satisfies the the gradient covariance matrices are positive definite:

$$V_{\tilde{\theta}}(\theta) = \mathrm{Cov}_{z\sim\mathcal{D}(\tilde{\theta})}\big(\nabla_\theta \ell(z;\theta)\big) \succ 0, \quad V_{\tilde{\theta}}^f(\theta) = \mathrm{Cov}_{x\sim\mathcal{D}_{\mathcal{X}}(\tilde{\theta})}(\nabla_\theta \ell(x, f(x);\theta)) \succ 0,$$

for any $\tilde{\theta}, \theta$.

(b). For any sample size n, assume $\hat{\theta}_t^{\mathrm{PPI}}$ has a density function with respect to the Lebesgue measure, and its characteristic function is absolutely integrable.

A.2 Details of Section 3: Theory of Inference under Performativity

We provide the omitted details in Section 3.

A.2.1 Consistency and Central Limit Theorem of $\hat{\theta}_t$

Let us denote:

$$\mathcal{L}_{\tilde{\theta}}(\theta) := \mathbb{E}_{z\sim\mathcal{D}(\tilde{\theta})}\ell(z;\theta), \quad \mathcal{L}_{\tilde{\theta},n}(\theta) := \frac{1}{n}\sum_{i=1}^n \ell(z_i;\theta),$$

where the samples $z_i = (x_i, y_i) \sim \mathcal{D}(\tilde{\theta})$ are drawn from the distribution under $\tilde{\theta}$.

Proposition A.6 (Consistency of $\hat{\theta}_t$, Restatement of Proposition 3.3)**.** *Under Assumption 3.1, if $\varepsilon < \frac{\gamma}{\beta}$, then for any given $T \geq 0$, we have that for all $t \in [T]$,*

$$\hat{\theta}_t \xrightarrow{P} \theta_t.$$

Proof. Let us denote $\hat{G}(\theta) := \mathrm{argmin}_{\theta' \in \Theta} \frac{1}{n}\sum_{i=1}^n \ell(z_i;\theta')$ where the samples $z_i \sim \mathcal{D}(\theta)$ are drawn for some parameter θ along the dynamic trajectory $\theta_0 \to \hat{\theta}_1 \to \cdots \hat{\theta}_t \to \cdots$.

$$\begin{aligned}
\|\theta_t - \hat{\theta}_t\| &= \|G(\theta_{t-1}) - \hat{G}(\hat{\theta}_{t-1})\| \\
&\leq \|G(\hat{\theta}_{t-1}) - \hat{G}(\hat{\theta}_{t-1})\| + \|G(\theta_{t-1}) - G(\hat{\theta}_{t-1})\| \\
&\leq \|G(\hat{\theta}_{t-1}) - \hat{G}(\hat{\theta}_{t-1})\| + \varepsilon\frac{\beta}{\gamma}\|\theta_{t-1} - \hat{\theta}_{t-1}\|,
\end{aligned}$$

where the last inequality follows from the results derived by [23], under Assumption 3.1, we have $\|G(\theta) - G(\theta')\| \leq \frac{\varepsilon\beta}{\gamma}\|\theta - \theta'\|$.

Notice that $\mathbb{E}(\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta)) = \mathcal{L}_{\hat{\theta}_{t-1}}(\theta)$. By local Lipschitz condition, there exists $\epsilon_0 > 0$ such that

$$\sup_{\theta:\|\theta - G(\hat{\theta}_{t-1})\| \leq \epsilon_0} |\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1}}(\theta)| \xrightarrow{P} 0.$$

Since ℓ is strongly convex for any θ, $G(\hat{\theta}_{t-1})$ is unique. Then we know that there exists δ such that $\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1}}(G(\hat{\theta}_{t-1})) > \delta$ for all θ in $\{\theta \mid \|\theta - G(\hat{\theta}_{t-1})\| = \epsilon_0\}$. Then it follows that:

$$\begin{aligned}
&\inf_{\|\theta - G(\hat{\theta}_{t-1})\| = \epsilon_0} \mathcal{L}_{\hat{\theta}_{t-1},n}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1},n}(G(\hat{\theta}_{t-1})) \\
=&\inf_{\|\theta - G(\hat{\theta}_{t-1})\| = \epsilon_0} \Big((\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1}}(\theta)) + (\mathcal{L}_{\hat{\theta}_{t-1}}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1}}(G(\hat{\theta}_{t-1}))) \\
&+ (\mathcal{L}_{\hat{\theta}_{t-1}}(G(\hat{\theta}_{t-1})) - \mathcal{L}_{\hat{\theta}_{t-1},n}(G(\hat{\theta}_{t-1})))\Big) \\
\geq& \delta - o_P(1).
\end{aligned}$$

Then we consider any fixed θ such that $\|\theta - G(\hat{\theta}_{t-1})\| \geq \epsilon_0$ it follows that

$$\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta) - \mathcal{L}_{\hat{\theta}_{t-1},n}(G(\hat{\theta}_{t-1})) \geq \frac{\theta - G(\hat{\theta}_{t-1})}{\omega - G(\hat{\theta}_{t-1})}\left(\mathcal{L}_{\hat{\theta}_{t-1},n}(\omega) - \mathcal{L}_{\hat{\theta}_{t-1},n}(G(\hat{\theta}_{t-1}))\right)$$

$$\geq \frac{\|\theta - G(\hat{\theta}_{t-1})\|}{\epsilon_0}(\delta - o_P(1)) \geq \delta - o_P(1),$$

where the first inequality holds for any ω by the convexity condition of $\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta)$, and the second inequality holds as we take $\omega = \frac{\theta - G(\hat{\theta}_{t-1})}{\|\theta - G(\hat{\theta}_{t-1})\|}\epsilon_0 + G(\hat{\theta}_{t-1})$ and using the above result. Thus no θ such that $\|\theta - G(\hat{\theta}_{t-1})\| = \epsilon_0$ can be the minimizer of $\mathcal{L}_{\hat{\theta}_{t-1},n}(\theta)$. Then $\|G(\hat{\theta}_{t-1}) - \hat{G}(\hat{\theta}_{t-1})\| \xrightarrow{P} 0$.

We then have, for a given $T \geq 0$, we have that for all $t \in [T]$,

$$\|\hat{\theta}_t - \theta_t\| \leq \sum_{i=0}^{t} (\varepsilon \frac{\beta}{\gamma})^{t-i} \|G(\hat{\theta}_i) - \hat{G}(\hat{\theta}_i)\| \xrightarrow{P} 0.$$

Thus, we conclude that $\hat{\theta}_t \xrightarrow{P} \theta_t$. □

Theorem A.7 (Central Limit Theorem of $\hat{\theta}_t$, Restatement of Theorem 3.4)**.** *Under Assumption 3.1 and A.1, if $\varepsilon < \frac{\gamma}{\beta}$, then for any given $T \geq 0$, we have that for all $t \in [T]$,*

$$\sqrt{n}(\hat{\theta}_t - \theta_t) \xrightarrow{D} \mathcal{N}(0, V_t)$$

with

$$V_t = \sum_{i=1}^{t} \left[\prod_{k=i}^{t-1} \nabla G(\theta_k)\right] \Sigma_{\theta_{i-1}}(\theta_i) \left[\prod_{k=i}^{t-1} \nabla G(\theta_k)\right]^\top .$$

In particular, $\nabla G(\theta_k) = -H_{\theta_k}(\theta_{k+1})^{-1}\left(\nabla_{\tilde{\theta}} \mathbb{E}_{z\sim\mathcal{D}(\theta_k)} \nabla_\theta \ell(z; \theta_{k+1})\right)$, where $\nabla_{\tilde{\theta}}$ is taking gradient for the parameter in $\mathcal{D}(\tilde{\theta})$, ∇_θ is taking gradient for the parameter in $\ell(z;\theta)$ and $\prod_{k=t}^{t-1} \nabla G(\theta_k) = I_d$.

Proof. Let $U_t := \sqrt{n}(\hat{\theta}_t - \theta_t)$ and denote $\tilde{\theta}_t = G(\hat{\theta}_{t-1})$. We make the following decomposition:

$$\hat{\theta}_t - \theta_t = \underbrace{(\tilde{\theta}_t - \theta_t)}_{(1)} + \underbrace{(\hat{\theta}_t - \tilde{\theta}_t)}_{(2)}.$$

Step 1: Conditional distribution of $U_t | U_{t-1}$.

For term (1), we have

$$\sqrt{n}(\tilde{\theta}_t - \theta_t) = \sqrt{n}(G(\hat{\theta}_{t-1}) - G(\theta_{t-1})).$$

For term (2), the empirical process analysis in [1] establishes that

$$\sqrt{n}(\hat{\theta}_t - \tilde{\theta}_t) \mid \hat{\theta}_{t-1} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)),$$

where the variance is given by

$$\Sigma_{\hat{\theta}_{t-1}}(\tilde{\theta}_t) = H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)^{-1} V_{\hat{\theta}_{t-1}}(\tilde{\theta}_t) H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)^{-1}.$$

Conditioning on $\hat{\theta}_{t-1}$ and considering the distribution $D(\hat{\theta}_{t-1})$, for any function h, we use the following shorthand notations:

$$\mathbb{E}_n h := \frac{1}{n}\sum_{i=1}^{n} h(x_i, y_i), \quad \mathbb{G}_n h := \sqrt{n}(\mathbb{E}_n h - \mathbb{E}_{(x,y)\sim\mathcal{D}(\hat{\theta}_{t-1})}[h(x,y)]).$$

Note that $\tilde{\theta}_t = G(\hat{\theta}_{t-1})$. Recall that

$$\mathcal{L}_{\tilde{\theta}}(\theta) := \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\tilde{\theta})} \ell(x, y; \theta), \quad \mathcal{L}_{\tilde{\theta},n} := \frac{1}{n}\sum_{i=1}^{n} \ell(x_i, y_i; \theta), \text{ where } (x_i, y_i) \sim \mathcal{D}(\tilde{\theta}).$$

Under the assumptions, Lemma 19.31 in [29] implies that for every sequence $h_n = O_P(1)$, we have

$$\mathbb{G}_n \left[\sqrt{n}\left(\ell(x, y; \tilde{\theta}_t + \frac{h_n}{\sqrt{n}}) - \ell(x, y; \tilde{\theta}_t)\right) - h_n^\top \nabla_\theta \ell(x, y; \tilde{\theta}_t)\right] \xrightarrow{P} 0.$$

Applying second-order Taylor expansion, we obtain that

$$\begin{aligned} n\mathbb{E}_n\left(\ell(x,y;\tilde{\theta}_t+\frac{h_n}{\sqrt{n}})-\ell(x,y;\tilde{\theta}_t)\right) &= n\left(\mathcal{L}_{\hat{\theta}_{t-1}}(\tilde{\theta}_t+\frac{h_n}{\sqrt{n}})-\mathcal{L}_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)\right) \\ &\quad + h_n^\top \mathbb{G}_n \nabla_\theta \ell(x,y;\tilde{\theta}_t)+o_p(1) \\ &= \frac{1}{2}h_n^\top H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)h_n + h_n^\top \mathbb{G}_n \nabla_\theta \ell(x,y;\tilde{\theta}_t)+o_p(1). \end{aligned}$$

Set $h_n^* = \sqrt{n}(\hat{\theta}_t-\tilde{\theta}_t)$ and $h_n = -H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)^{-1}\mathbb{G}_n\nabla_\theta\ell(x,y;\tilde{\theta}_t)$, Corollary 5.53 in [29] implies they are $O_P(1)$.

Since $\hat{\theta}_t$ is the minimizer of $\mathcal{L}_{n,\hat{\theta}_{t-1}}$, the first term is smaller than the second term. We can rearrange the terms and obtain:

$$\frac{1}{2}(h_n^*-h_n)^T H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)(h_n^*-h_n) = o_P(1),$$

which leads to $h_n^*-h_n = O_P(1)$. Then the above asymptotic normality result follows directly by applying the central limit theorem (CLT) to the following terms, conditioning on $\hat{\theta}_{t-1}$:

$$\begin{aligned} &\sqrt{n}(\hat{\theta}_t-\tilde{\theta}_t)\mid\hat{\theta}_{t-1} = -H_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)^{-1}S+o_P(1), \\ &S=\sqrt{\frac{1}{n}}\sum_{i=1}^n\Big(\nabla_\theta\ell(x_{t,i},y_{t,i};\tilde{\theta}_t)-\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\hat{\theta}_{t-1})}[\nabla_\theta\ell(x,y;\tilde{\theta}_t)]\Big). \end{aligned}$$

Note that, conditioning on $\hat{\theta}_{t-1}$, (1) is a constant. Therefore, (1) and (2) follow a joint Gaussian distribution. Consequently, given U_{t-1}, the conditional distribution of U_t is given by:

$$\begin{aligned} U_t\mid U_{t-1} &= \sqrt{n}(\hat{\theta}_t-\theta_t)\mid\hat{\theta}_{t-1} \\ &= \sqrt{n}(\tilde{\theta}_t-\theta_t)+\sqrt{n}(\hat{\theta}_t-\tilde{\theta}_t)\mid\hat{\theta}_{t-1} \\ &= \sqrt{n}(G(\hat{\theta}_{t-1})-G(\theta_{t-1}))+\sqrt{n}(\hat{\theta}_t-\tilde{\theta}_t)\mid\hat{\theta}_{t-1} \\ &\xrightarrow{D}\mathcal{N}\left(\sqrt{n}(G(\hat{\theta}_{t-1})-G(\theta_{t-1})),\Sigma_{\hat{\theta}_{t-1}}(\tilde{\theta}_t)\right). \\ &= \mathcal{N}\left(\sqrt{n}(G(\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1})-G(\theta_{t-1})),\Sigma_{\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1}}(G(\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1}))\right). \end{aligned}$$

For later references, we denote $U_t\mid U_{t-1}\xrightarrow{D}\mathcal{N}(\mu(U_{t-1}),\Sigma(U_{t-1}))$.

Step 2: Marginal distribution of U_t. We calculate the characteristic function of U_t by induction. To begin with, we directly have

$$X_1\xrightarrow{D}\mathcal{N}(0,V_1),\quad V_1=\Sigma_{\theta_0}(\theta_1).$$

Now, assume that $U_{t-1}\xrightarrow{D}\mathcal{N}(0,V_{t-1})$, we derive the joint distribution of (U_t,U_{t-1}) and marginal distribution of U_t. Then we have, the characteristics functions ϕ and the probability density function p of distributions U_{t-1} and $U_t\mid U_{t-1}$ follow:

$$\phi_{U_{t-1}}(s)\to\phi_{\mathcal{N}(0,V_{t-1})}(s)=\exp(-\frac{1}{2}s^TV_{t-1}s),\quad p_{U_{t-1}}(u)=\frac{1}{(2\pi)^d}\int e^{-iz^Tu}\phi_{U_{t-1}}(z)dz,$$

$$\phi_{U_t|U_{t-1}}(s)\to\phi_{\mathcal{N}(\mu(U_{t-1}),\Sigma(U_{t-1}))}(s)=\exp(is^T\mu(U_{t-1})-\frac{1}{2}s^T\Sigma(U_{t-1})s).$$

Then we have

$$\begin{aligned} &\phi_{U_t}(s)=\mathbb{E}e^{is^TU_t}=\mathbb{E}\big(\mathbb{E}(e^{is^\top U_t}\mid U_{t-1})\big)=E_{U_{t-1}}\phi_{U_t|U_{t-1}}(s\mid U_{t-1}) \\ &=\int\phi_{U_t|U_{t-1}}(s\mid u)\,p_{U_{t-1}}(u)\,du \\ &=\int\phi_{U_t|U_{t-1}}(s\mid u)\,\frac{1}{(2\pi)^d}\int e^{-iz^\top u}\,\phi_{U_{t-1}}(z)\,dz\,du \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(2\pi)^d}\iint \phi_{U_t|U_{t-1}}(s \mid u)\,\phi_{U_{t-1}}(z)\,e^{-iz^\top u}\,dz\,du\\
&= \frac{1}{(2\pi)^d}\iint \exp\big(is^\top\mu(U_{t-1}) - \tfrac{1}{2}\,s^\top\Sigma(U_{t-1})s\big)\,\exp\big(-\tfrac{1}{2}\,z^\top V_{t-1}z\big)\,e^{-iz^\top u}\,dz\,du\\
&= \frac{1}{(2\pi)^d}\int \exp\big(is^\top\mu(U_{t-1}) - \tfrac{1}{2}\,s^\top\Sigma(U_{t-1})s\big)\Big(\int \exp\big(-\tfrac{1}{2}\,z^\top V_{t-1}z - i\,z^\top u\big)\,dz\Big)\,du\\
&= \frac{1}{(2\pi)^d}\int \exp\big(is^\top\mu(U_{t-1}) - \tfrac{1}{2}\,s^\top\Sigma(U_{t-1})s\big)\\
&\times\Big(\int \exp\big(-\tfrac{1}{2}\,u^\top V_{t-1}^{-1}u\big)\exp\big(-\tfrac{1}{2}\,(z - V_{t-1}^{-1}iu)^\top V_{t-1}(z - V_{t-1}^{-1}iu)\big)\,dz\Big)du\\
&= \frac{1}{(2\pi)^d}\int \exp\big(is^\top\mu(U_{t-1}) - \tfrac{1}{2}\,s^\top\Sigma(U_{t-1})s\big)((2\pi)^{\frac{d}{2}}\,\frac{1}{\det|V_{t-1}|}\cdot\exp(-\frac{1}{2}u^\top V_{t-1}^{-1}u))\,du\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\det|V_{t-1}|}\int \exp(is^T\mu(U_{t-1}) - \frac{1}{2}s^T\Sigma(U_{t-1})s - \frac{1}{2}u^T V_{t-1}u)du.
\end{aligned}
$$

Apply dominant convergence theorem to $\lim_{n\to\infty}\phi_{U_t}(s)$, we have:

$$
\begin{aligned}
&\lim_{n\to\infty}\phi_{U_t}(s) = \lim_{n\to\infty}\frac{1}{(2\pi)^{\frac{d}{2}}\det|V_{t-1}|}\int \exp(is^T\mu(U_{t-1}) - \frac{1}{2}s^T\Sigma(U_{t-1})s - \frac{1}{2}u^T V_{t-1}u)du\\
&= \lim_{n\to\infty}\frac{1}{(2\pi)^{\frac{d}{2}}\det|V_{t-1}|}\int \exp(is^T\sqrt{n}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1}))\\
&- \frac{1}{2}s^T\Sigma_{\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1}}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}))s - \frac{1}{2}u^T V_{t-1}u)du\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\det|V_{t-1}|}\int \lim_{n\to\infty}\exp(is^T\sqrt{n}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1}))\\
&- \frac{1}{2}s^T\Sigma_{\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1}}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}))s - \frac{1}{2}u^T V_{t-1}u)du\\
&= \frac{1}{(2\pi)^{\frac{d}{2}}\det|V_{t-1}|}\int \exp(is^T\nabla G(\theta_{t-1})u - \frac{1}{2}s^T\Sigma_{\theta_{t-1}}(G(\theta_{t-1}))s - \frac{1}{2}u^T V_{t-1}u)du\\
&= \exp(-\frac{1}{2}s^T\nabla G(\theta_{t-1})V_{t+1}\nabla G(\theta_{t-1})^T s - \frac{1}{2}s^T\Sigma_{U_{t-1}}(\theta_t)s),
\end{aligned}
$$

which is the characteristic function of $\mathcal{N}(0, V_t)$, where $V_t = \nabla G(\theta_{t-1})V_{t-1}\nabla G(\theta_{t-1})^\top + \Sigma_{\theta_{t-1}}(\theta_t)$. Here we use the fact that $\lim_{n\to\infty}\sqrt{n}\Big(G(\frac{y}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1})\Big) = \nabla G(\theta_{t-1})y$, and the dominant convergence theorem holds as we have

$$
\begin{aligned}
&|\exp(is^T\sqrt{n}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1}))\\
&- \frac{1}{2}s^T\Sigma_{\frac{U_{t-1}}{\sqrt{n}}+\theta_{t-1}}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}))s - \frac{1}{2}u^T V_{t-1}u)| \le |\exp(-\frac{1}{2}u^T V_{t-1}u)|.
\end{aligned}
$$

Thus we conclude by induction that

$$
\begin{aligned}
&U_t \xrightarrow{D} \mathcal{N}\left(0, V_t\right),\\
&V_t = \sum_{i=1}^{t}\left[\prod_{k=i}^{t-1}\nabla G(\theta_k)\right]\Sigma_{\theta_{i-1}}(\theta_i)\left[\prod_{k=i}^{t-1}\nabla G(\theta_k)\right]^\top.
\end{aligned}
$$

And by the theorem of implicit function, we can calculate the gradient of G as follows

$$
\nabla G(\theta) = -\left[\big(\nabla_\psi^2 \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\theta)}\ell(x,y;\psi)\big)|_{\psi=G(\theta)}\right]^{-1}\big(\nabla_\psi\nabla_{\tilde{\theta}}\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\theta)}\ell(x,y;\psi)\big)|_{\psi=G(\theta)}.
$$

$$\begin{aligned}\nabla G(\theta_k) &= -\left[\mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\theta_k)} \nabla_\theta^2 \ell(x,y;\theta_{k+1})\right]^{-1} \big(\nabla_{\tilde{\theta}} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\theta_k)} \nabla_\theta \ell(x,y;\theta_{k+1})\big) \\ &= -H_{\theta_k}(\theta_{k+1})^{-1}\big(\nabla_{\tilde{\theta}} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\theta_k)} \nabla_\theta \ell(x,y;\theta_{k+1})\big) \\ &= -H_{\theta_k}(\theta_{k+1})^{-1}\mathbb{E}_{z\sim\mathcal{D}(\theta_k)}[\nabla_\theta \ell(z,\theta_{k+1})\nabla_\theta \log p(z,\theta_k)^\top].\end{aligned}$$

□

A.2.2 Score Matching

In this part, we provide details about our score matching mechanism.

Given that

$$\nabla G(\theta_k) = -H_{\theta_k}(\theta_{k+1})^{-1}\mathbb{E}_{z\sim\mathcal{D}(\theta_k)}[\nabla_\theta \ell(z,\theta_{k+1})\nabla_\theta \log p(z,\theta_k)^\top],$$

once we have a good estimation of $\nabla_\theta \log p(z,\theta_k)$ for all $z \in \mathcal{Z}$, $\nabla G(\theta_k)$ could be easily estimated by samples.

Recall that we use a model $M(z,\theta;\psi)$ parameterized by ψ to approximate $p(z,\theta)$. Inspired by the objective in [12], for any given θ (e.g., $\hat{\theta}_t$), we aim to optimize the following objective parameterized by ψ:

$$\begin{aligned}J(\psi) &= \int p(z,\theta)\|\nabla_\theta \log p(z,\theta) - s(z,\theta;\psi)\|^2 dz \\ &= \int p(z,\theta)\Big(\|\nabla_\theta \log p(z,\theta)\|^2 + \|s(z,\theta;\psi)\|^2 - 2\nabla_\theta \log p(z,\theta)^\top s(z,\theta;\psi)\Big)dz\end{aligned}$$

where $s(z,\theta;\psi) = \nabla_\theta \log M(z,\theta;\psi)$.

As mentioned in the main context, the first term is unrelated to ψ; the second term involves model M that is chosen by us, so we have the analytical expression of $s(z,\theta;\psi)$. Thus, our key task will be estimating the third term, which involves $\mathcal{K}(z,\theta;\psi) := \int p(z,\theta)\nabla_\theta \log p(z,\theta)^\top s(z,\theta;\psi)dz$.

Lemma A.8 (Restatement of lemma 3.5)**.** *Under Assumption A.2, we have*

$$\mathcal{K}(z,\theta;\psi) = \sum_{i=1}^d \left[\frac{\partial}{\partial\theta^{(i)}}\int p(z,\theta)\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz - \int p(z,\theta)\frac{\partial^2 \log M(z,\theta;\psi)}{\partial\theta^{(i)2}}dz\right]$$

where $\theta^{(i)}$ is the i-th coordinate of θ.

Proof. Recall that θ is of d-dimension.

$$\begin{aligned}\int p(z,\theta)\nabla_\theta \log p(z,\theta)^\top s(z,\theta;\psi)dz &= \sum_{i=1}^d \int p(z,\theta)\frac{\partial \log p(z,\theta)}{\partial\theta^{(i)}}\cdot\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz \\ &= \sum_{i=1}^d \int p(z,\theta)\frac{\partial \log p(z,\theta)}{\partial\theta^{(i)}}\cdot\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz \\ &= \sum_{i=1}^d \int \frac{\partial p(z,\theta)}{\partial\theta^{(i)}}\cdot\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz.\end{aligned}$$

Then, we study $\int \frac{\partial p(z,\theta)}{\partial\theta^{(i)}}\cdot\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz$. Under Assumption A.2, the integral and differentiation of the following equation is exchangeable, i.e.,

$$\frac{\partial}{\partial\theta^{(i)}}\int p(z,\theta)\frac{\partial M(z,\theta;\psi)}{\partial\theta^{(i)}}dz = \int \frac{p(z,\theta)}{\partial\theta^{(i)}}\frac{\partial M(z,\theta;\psi)}{\partial\theta^{(i)}}dz.$$

According to integral by parts, we have

$$\int \frac{\partial p(z,\theta)}{\partial\theta^{(i)}}\cdot\frac{\partial \log M(z,\theta;\psi)}{\partial\theta^{(i)}}dz = \frac{\partial}{\partial\theta^{(i)}}\int p(z,\theta)\frac{\partial M(z,\theta;\psi)}{\partial\theta^{(i)}}dz - \int p(z,\theta)\frac{\partial^2 \log M(z,\theta;\psi)}{\partial\theta^{(i)2}}dz.$$

Thus, our proof is completed. □

The rest of the estimation process via policy perturbation is provided in the main context in Section 3.

The other part omitted in Section 3 is the details about Eq. 2 that

$$\hat{V}_t^{-1/2}\sqrt{n}(\hat{\theta}_t - \theta_t) \xrightarrow{D} \mathcal{N}(0, I_d).$$

Here $\hat{V}_t$ denotes the sample-based estimator of the variance, obtained by plugging in the empirical Hessian and empirical covariance matrices:

$$\widehat{H}_{\hat{\theta}_{t-1}}(\hat{\theta}_t) = \widehat{\mathbb{E}}_{z\sim\mathcal{D}(\hat{\theta}_{t-1})}\nabla^2_\theta \ell(z;\hat{\theta}_t), \quad \widehat{\mathrm{Cov}}_{z\sim\mathcal{D}(\hat{\theta}_{t-1})}\big(\nabla_\theta \ell(z;\hat{\theta}_t)\big),$$

as well as the estimator for $\nabla G(\hat{\theta}_{t-1})$:

$$-\widehat{H}_{\hat{\theta}_{t-1}}(\hat{\theta}_t)^{-1}\widehat{\mathbb{E}}_{z\sim\mathcal{D}(\hat{\theta}_{t-1})}[\nabla_\theta \ell(z,\hat{\theta}_t)\nabla_\theta \log M(z,\hat{\theta}_{t-1},\hat{\psi})^\top],$$

where $\hat{\psi}$ is obtained by minimizing $\hat{J}_{n,k}$.

Eq. 2 is a direct result following Slutsky's theorem. Assumption A.3 makes sure the empirical optimizer set can converge to the population optimizer set. Then other parts such as estimation of the Hessian matrix etc. could all be directly obtained by standard law of large numbers. Thus, we can directly use Slutsky's theorem to obtain Eq. 2.

A.3 Details of Section 4: Theory of Prediction-Powered Inference under Performativity

Without loss of generality, we let $N_t = N$ and $n_t = n$ for all $t \in [T]$.

Let us denote

$$\mathcal{L}_{\tilde{\theta}}(\theta) := \mathbb{E}_{(x,y)\sim\mathcal{D}(\tilde{\theta})}\ell(x,y;\theta), \quad \mathcal{L}^{f,\lambda}_{\tilde{\theta}}(\theta) := \mathcal{L}_{\tilde{\theta},n}(\theta) + \lambda\cdot(\widetilde{\mathcal{L}}^f_{\tilde{\theta},N}(\theta) - \mathcal{L}^f_{\tilde{\theta},n}(\theta)),$$

where

$$\mathcal{L}_{\tilde{\theta},n}(\theta) := \frac{1}{n}\sum_{i=1}^n \ell(x_i,y_i;\theta),\ \mathcal{L}^f_{\tilde{\theta},n}(\theta) := \frac{1}{n}\sum_{i=1}^n \ell\left((x_i, f(x_i));\theta\right),\ \widetilde{\mathcal{L}}^f_{\tilde{\theta},N}(\theta) := \frac{1}{N}\sum_{i=1}^N \ell\left((x^u_i, f(x^u_i));\theta\right).$$

Here the samples $(x_i, y_i) \sim \mathcal{D}(\tilde{\theta})$ and $x^u_i \sim \mathcal{D}_{\mathcal{X}}(\tilde{\theta})$ are drawn from the distribution under $\tilde{\theta}$. Recall that we have defined $\Sigma_{\lambda,\tilde{\theta}}(\theta) = H_{\tilde{\theta}}(\theta)^{-1}\left(rV^f_{\lambda,\tilde{\theta}}(\theta) + V_{\lambda,\tilde{\theta}}(\theta)\right)H_{\tilde{\theta}}(\theta)^{-1}$ before Theorem 4.1 (in the following we sometimes omit r for simplicity).

Theorem A.9 (Consistency of $\hat{\theta}^{\mathrm{PPI}}_t$)**.** *Under Assumption 3.1 and A.4, if $\varepsilon < \frac{\gamma}{\beta}$, then for any given $T \geq 0$, we have that for all $t \in [T]$,*

$$\hat{\theta}^{PPI}_{t+1}(\lambda_t) \xrightarrow{P} \theta_{t+1}.$$

Proof. Let us denote $\hat{G}^f_\lambda(\theta) := \mathrm{argmin}_{\theta'\in\Theta}\frac{\lambda}{N}\sum_{i=1}^N \ell(x^u_i, f(x^u_i);\theta') + \frac{1}{n}\sum_{i=1}^n\big(\ell(x_i,y_i;\theta') - \lambda\ell(x_i, f(x_i);\theta')\big)$, where the samples $(x_i,y_i)\sim\mathcal{D}(\theta)$ and $x^u_i \sim \mathcal{D}_{\mathcal{X}}(\theta)$ are drawn for some parameter θ along the dynamic trajectory $\theta_0 \to \hat{\theta}_1 \to \cdots \hat{\theta}_t \to \cdots$.

$$\begin{aligned}\|\theta_t - \hat{\theta}^{\mathrm{PPI}}_t\| &= \|G(\theta_{t-1}) - \hat{G}^f_{\lambda_t}(\hat{\theta}^{\mathrm{PPI}}_{t-1})\| \\ &\leq \|G(\hat{\theta}^{\mathrm{PPI}}_{t-1}) - \hat{G}^f_{\lambda_t}(\hat{\theta}^{\mathrm{PPI}}_{t-1})\| + \|G(\theta_{t-1}) - G(\hat{\theta}^{\mathrm{PPI}}_{t-1})\| \\ &\leq \|G(\hat{\theta}^{\mathrm{PPI}}_{t-1}) - \hat{G}^f_{\lambda_t}(\hat{\theta}^{\mathrm{PPI}}_{t-1})\| + \varepsilon\frac{\beta}{\gamma}\|\theta_{t-1} - \hat{\theta}^{\mathrm{PPI}}_{t-1}\|,\end{aligned}$$

where the last inequality follows from the results derived by [23], under Assumption 3.1, we have $\|G(\theta) - G(\theta')\| \leq \frac{\varepsilon\beta}{\gamma}\|\theta - \theta'\|$.

Notice that $\mathbb{E}(\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\mathrm{PPI}}_{t-1}}(\theta)) = \mathcal{L}_{\hat{\theta}^{\mathrm{PPI}}_{t-1}}(\theta)$. By local Lipschitz condition, there exists $\epsilon_0 > 0$ such that

$$\sup_{\theta:\|\theta - G(\hat{\theta}^{\mathrm{PPI}}_{t-1})\|\leq\epsilon_0} |\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\mathrm{PPI}}_{t-1}}(\theta) - \mathcal{L}_{\hat{\theta}^{\mathrm{PPI}}_{t-1}}(\theta)| \xrightarrow{P} 0.$$

Since ℓ is strongly convex for any θ, $G(\hat{\theta}^{\text{PPI}}_{t-1})$ is unique. Then we know that there exists δ such that $\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta) - \mathcal{L}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})) > \delta$ for all θ in $\{\theta \mid \|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\| = \epsilon_0\}$. Then it follows that:

$$
\begin{aligned}
&\inf_{\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\| = \epsilon_0} \mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta) - \mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1}))\\
=& \inf_{\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\| = \epsilon_0} \Big((\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta) - \mathcal{L}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta)) + (\mathcal{L}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta) - \mathcal{L}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})))\\
&+ (\mathcal{L}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})) - \mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})))\Big)\\
\geq& \delta - o_P(1).
\end{aligned}
$$

Then we consider any fixed θ such that $\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\| \geq \epsilon_0$ it follows that

$$
\begin{aligned}
\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta) - \mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})) &\geq \frac{\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})}{\omega - G(\hat{\theta}^{\text{PPI}}_{t-1})} \left(\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\omega) - \mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(G(\hat{\theta}^{\text{PPI}}_{t-1})) \right)\\
&\geq \frac{\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\|}{\epsilon_0}(\delta - o_P(1)) \geq \delta - o_P(1),
\end{aligned}
$$

where the first inequality holds for any ω by the convexity condition of $\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta)$, and the second inequality holds as we take $\omega = \frac{\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})}{\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\|}\epsilon_0 + G(\hat{\theta}^{\text{PPI}}_{t-1})$ and using the above result. Thus no θ such that $\|\theta - G(\hat{\theta}^{\text{PPI}}_{t-1})\| = \epsilon_0$ can be the minimizer of $\mathcal{L}^{f,\lambda_t}_{\hat{\theta}^{\text{PPI}}_{t-1}}(\theta)$. Then $\|G(\hat{\theta}^{\text{PPI}}_{t-1}) - \hat{G}^f_{\lambda_t}(\hat{\theta}^{\text{PPI}}_{t-1})\| \xrightarrow{P} 0$.

We then have, for a given $T \geq 0$, we have that for all $t \in [T]$,

$$
\|\hat{\theta}^{\text{PPI}}_t - \theta_t\| \leq \sum_{i=0}^{t} (\varepsilon \frac{\beta}{\gamma})^{t-i} \|G(\hat{\theta}^{\text{PPI}}_i) - \hat{G}^f_{\lambda_i}(\hat{\theta}^{\text{PPI}}_i)\| \xrightarrow{P} 0.
$$

Thus, we conclude that $\hat{\theta}^{\text{PPI}}_t \xrightarrow{P} \theta_t$. $\square$

Theorem A.10 (Central Limit Theorem of $\hat{\theta}^{\text{PPI}}_t(\lambda_t)$, Restatement of Theorem 4.1)**.** *Under Assumption 3.1, A.4, and A.5, if $\varepsilon < \frac{\gamma}{\beta}$ and $\frac{n}{N} \to r$ for some $r \geq 0$, then for any given $T \geq 0$, we have that for all $t \in [T]$,*

$$
\sqrt{n}\big(\hat{\theta}^{PPI}_t(\lambda_t) - \theta_t\big) \xrightarrow{D} \mathcal{N}\Big(0, V^{PPI}_t\big(\{\lambda_j, \theta_j\}_{j=1}^t; r\big)\Big)
$$

with

$$
V^{PPI}_t\big(\{\lambda_j, \theta_j\}_{j=1}^t; r\big) = \sum_{i=1}^{t} \left[\prod_{k=i}^{t-1} \nabla G(\theta_k)\right] \Sigma_{\lambda_i, \theta_{i-1}}(\theta_i; r) \left[\prod_{k=i}^{t-1} \nabla G(\theta_k)\right]^\top .
$$

Proof. Let us denote the variance terms by V^{PPI}_t for simplicity, while omitting explicit dependence on parameters in the notation. Let $U_t := \sqrt{n}(\hat{\theta}^{\text{PPI}}_t - \theta_t)$ and denote $\tilde{\theta}_t = G(\hat{\theta}^{\text{PPI}}_{t-1})$. We make the following decomposition:

$$
\hat{\theta}^{\text{PPI}}_t - \theta_t = \underbrace{(\tilde{\theta}_t - \theta_t)}_{(1)} + \underbrace{(\hat{\theta}^{\text{PPI}}_t - \tilde{\theta}_t)}_{(2)}.
$$

Step 1: Conditional distribution of $U_t | U_{t-1}$.

For term (1), we have

$$
\sqrt{n}(\tilde{\theta}_t - \theta_t) = \sqrt{n}(G(\hat{\theta}^{\text{PPI}}_{t-1}) - G(\theta_{t-1})).
$$

For term (2), the empirical process analysis in [1] establishes that

$$
\sqrt{n}(\hat{\theta}^{\text{PPI}}_t - \tilde{\theta}_t) | \hat{\theta}^{\text{PPI}}_{t-1} \xrightarrow{D} \mathcal{N}(0, \Sigma_{\lambda_t, \hat{\theta}^{\text{PPI}}_{t-1}}(\tilde{\theta}_t; r)),
$$

where the variance is given by

$$\Sigma_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t; r) = H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)^{-1}\left(rV^f_{\lambda_t,\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t) + V_{\lambda_t,\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)\right)H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)^{-1}.$$

Conditioning on $\hat\theta^{\text{PPI}}_{t-1}$, for any function h, we use the following shorthand notations:

$$\begin{aligned}
&\mathbb{E}_n h := \frac{1}{n}\sum_{i=1}^n h(x_i, y_i), \quad \mathbb{G}_n h := \sqrt{n}(\mathbb{E}_n h - \mathbb{E}_{(x,y)\sim\mathcal{D}(\hat\theta^{\text{PPI}}_{t-1})}[h(x,y)]),\\
&\hat{\mathbb{E}}^f_N h := \frac{1}{N}\sum_{i=1}^N h(x^u_i, f(x^u_i)), \quad \hat{\mathbb{G}}^f_N h := \sqrt{N}(\hat{\mathbb{E}}_N h - \mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}(\hat\theta^{\text{PPI}}_{t-1})}[h(x, f(x))]),\\
&\hat{\mathbb{E}}^f_n h := \frac{1}{n}\sum_{i=1}^n h(x_i, f(x_i)), \quad \hat{\mathbb{G}}^f_n h := \sqrt{n}(\hat{\mathbb{E}}_n h - \mathbb{E}_{x\sim\mathcal{D}_{\mathcal{X}}(\hat\theta^{\text{PPI}}_{t-1})}[h(x, f(x))]).
\end{aligned}$$

Note that $\tilde\theta_t = G(\hat\theta^{\text{PPI}}_{t-1})$. Recall that

$$\mathcal{L}_{\tilde\theta}(\theta) := \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}(\tilde\theta)} \ell(x, y; \theta), \quad \mathcal{L}^{f,\lambda}_{\tilde\theta}(\theta) := \mathcal{L}_{\tilde\theta,n}(\theta) + \lambda\cdot(\widetilde{\mathcal{L}}^f_{\tilde\theta,N}(\theta) - \mathcal{L}^f_{\tilde\theta,n}(\theta)).$$

Under the assumptions, Lemma 19.31 in [29] implies that for every sequence $h_n = O_P(1)$, we have

$$\begin{aligned}
&\mathbb{G}_n\left[\sqrt{n}\left(\ell(x,y;\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \ell(x,y;\tilde\theta_t)\right) - h_n^\top\nabla_\theta\ell(x,y;\tilde\theta_t)\right] \xrightarrow{P} 0,\\
&\hat{\mathbb{G}}^f_N\left[\sqrt{n}\left(\ell(x,y;\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \ell(x,y;\tilde\theta_t)\right) - h_n^\top\nabla_\theta\ell(x,y;\tilde\theta_t)\right] \xrightarrow{P} 0,\\
&\hat{\mathbb{G}}^f_n\left[\sqrt{n}\left(\ell(x,y;\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \ell(x,y;\tilde\theta_t)\right) - h_n^\top\nabla_\theta\ell(x,y;\tilde\theta_t)\right] \xrightarrow{P} 0.
\end{aligned}$$

Applying second-order Taylor expansion, we obtain that

$$\begin{aligned}
n\mathbb{E}_n\left(\ell(x,y;\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \ell(x,y;\tilde\theta_t)\right) &= n\left(\mathcal{L}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \mathcal{L}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)\right) + h_n^\top\mathbb{G}_n\nabla_\theta\ell(x,y;\tilde\theta_t) + o_p(1)\\
&= \frac{1}{2}h_n^\top H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)h_n + h_n^\top\mathbb{G}_n\nabla_\theta\ell(x,y;\tilde\theta_t) + o_p(1).
\end{aligned}$$

Based on similar calculation of the previous two terms, we can obtain that:

$$\begin{aligned}
&n\left(\mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)\right)\\
=&\frac{1}{2}h_n^\top H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)h_n + h_n^\top\left(\mathbb{G}_n + \lambda\sqrt{\frac{n}{N}}\hat{\mathbb{G}}^f_N - \lambda\hat{\mathbb{G}}^f_n\right)\nabla_\theta\ell(x,y;\tilde\theta_t) + o_p(1).
\end{aligned}$$

By considering $h^*_n = \sqrt{n}(\hat\theta^{\text{PPI}}_t - \tilde\theta_t)$ and $h_n = -H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)^{-1}\left(\mathbb{G}_n + \lambda\sqrt{\frac{n}{N}}\hat{\mathbb{G}}^f_N - \lambda\hat{\mathbb{G}}^f_n\right)\nabla_\theta\ell(x,y;\tilde\theta_t)$, Corollary 5.53 in [29] implies they are $O_P(1)$ and we obtain that

$$\begin{aligned}
&n\left(\mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\hat\theta^{\text{PPI}}_t) - \mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)\right) = \frac{1}{2}h^{*\top}_n H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)h^*_n + h^{*\top}_n\left(\mathbb{G}_n + \lambda\sqrt{\frac{n}{N}}\hat{\mathbb{G}}^f_N - \lambda\hat{\mathbb{G}}^f_n\right)\nabla_\theta\ell(x,y;\tilde\theta_t) + o_p(1)\\
&n\left(\mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t + \frac{h_n}{\sqrt{n}}) - \mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)\right) = -\frac{1}{2}h_n^\top H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)h_n + o_P(1).
\end{aligned}$$

Since $\hat\theta^{\text{PPI}}_t$ is the minimizer of $\mathcal{L}^{f,\lambda}_{\hat\theta^{\text{PPI}}_{t-1}}$, the first term is smaller than the second term. We can rearrange the terms and obtain:

$$\frac{1}{2}(h^*_n - h_n)^T H_{\hat\theta^{\text{PPI}}_{t-1}}(\tilde\theta_t)(h^*_n - h_n) = o_P(1),$$

which leads to $h_n^* - h_n = O_P(1)$. Then the above asymptotic normality result follows directly by applying the central limit theorem (CLT) to the following terms, conditioning on $\hat{\theta}_{t-1}^{\text{PPI}}$:

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_t^{\text{PPI}} - \tilde{\theta}_t)|\hat{\theta}_{t-1}^{\text{PPI}} = -H_{\hat{\theta}_{t-1}^{\text{PPI}}}(\tilde{\theta}_t)^{-1}(S_1 + S_2) + o_P(1),\\
&S_1 = \lambda_t \sqrt{\frac{n}{N}} \sqrt{\frac{1}{N}} \sum_{i=1}^{N} \Big(\nabla_\theta \ell(x_{t,i}^u, f(x_{t,i}^u); \tilde{\theta}_t) - \mathop{\mathbb{E}}_{x \sim \mathcal{D}_{\mathcal{X}}(\hat{\theta}_{t-1}^{\text{PPI}})} \nabla_\theta \ell(x, f(x); \tilde{\theta}_t) \Big),\\
&S_2 = \sqrt{\frac{1}{n}} \sum_{i=1}^{n} \Big(\nabla_\theta \ell(x_{t,i}, y_{t,i}; \tilde{\theta}_t) - \lambda_t \nabla_\theta \ell(x_{t,i}, f(x_{t,i}); \tilde{\theta}_t)\\
&\quad - \mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}(\hat{\theta}_{t-1}^{\text{PPI}})} [\nabla_\theta \ell(x, y; \tilde{\theta}_t) - \lambda_t \nabla_\theta \ell(x, f(x); \tilde{\theta}_t)] \Big).
\end{aligned}
$$

Note that, conditioning on $\hat{\theta}_{t-1}^{\text{PPI}}$, (1) is a constant. Therefore, (1) and (2) follow a joint Gaussian distribution. Consequently, given U_{t-1}, the conditional distribution of U_t is given by:

$$
\begin{aligned}
U_t|U_{t-1} &= \sqrt{n}(\hat{\theta}_t^{\text{PPI}} - \theta_t)|\hat{\theta}_{t-1}^{\text{PPI}}\\
&= \sqrt{n}(\tilde{\theta}_t - \theta_t) + \sqrt{n}(\hat{\theta}_t^{\text{PPI}} - \tilde{\theta}_t)|\hat{\theta}_{t-1}^{\text{PPI}}\\
&= \sqrt{n}(G(\hat{\theta}_{t-1}^{\text{PPI}}) - G(\theta_{t-1})) + \sqrt{n}(\hat{\theta}_t^{\text{PPI}} - \tilde{\theta}_t)|\hat{\theta}_{t-1}^{\text{PPI}}\\
&\xrightarrow{D} \mathcal{N}\left(\sqrt{n}(G(\hat{\theta}_{t-1}^{\text{PPI}}) - G(\theta_{t-1})), \Sigma_{\lambda_t, \hat{\theta}_{t-1}^{\text{PPI}}}(\tilde{\theta}_t; r) \right).\\
&= \mathcal{N}\left(\sqrt{n}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1})), \Sigma_{\lambda_t, \frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}}(G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}); r) \right).
\end{aligned}
$$

For later references, we denote $U_t \mid U_{t-1} \xrightarrow{D} \mathcal{N}(\mu(U_{t-1}), \Sigma(U_{t-1}; r))$.

Step 2: Marginal distribution of U_t. We calculate the characteristic function of U_t by induction. To begin with, we directly have

$$
X_1 \xrightarrow{D} \mathcal{N}(0, V_1^{\text{PPI}}), \quad V_1^{\text{PPI}} = \Sigma_{\lambda_0, \theta_0}(\theta_1; r).
$$

Now, assume that $U_{t-1} \xrightarrow{D} \mathcal{N}(0, V_{t-1}^{\text{PPI}})$, we derive the joint distribution of (U_t, U_{t-1}) and marginal distribution of U_t.Then we have, the characteristics functions ϕ and the probability density function p of distributions U_{t-1} and $U_t \mid U_{t-1}$ follow:

$$
\phi_{U_{t-1}}(s) \to \phi_{\mathcal{N}(0, V_{t-1}^{\text{PPI}})}(s) = \exp(-\frac{1}{2} s^T V_{t-1}^{\text{PPI}} s), \quad p_{U_{t-1}}(u) = \frac{1}{(2\pi)^d} \int e^{-iz^T u} \phi_{U_{t-1}}(z) dz
$$

$$
\phi_{U_t|U_{t-1}}(s) \to \phi_{\mathcal{N}(\mu(U_{t-1}), \Sigma(U_{t-1}; r))}(s) = \exp(is^T \mu(U_{t-1}) - \frac{1}{2} s^T \Sigma(U_{t-1}; r) s).
$$

Then according to the proof of vanilla CLT under performativity in Section A.2, we have:

$$
\phi_{U_t}(s) = \frac{1}{(2\pi)^{\frac{d}{2}} \det |V_{t-1}^{\text{PPI}}|} \int \exp\left(is^T \mu(U_{t-1}) - \frac{1}{2} s^T \Sigma(U_{t-1}; r) s - \frac{1}{2} u^T V_{t-1}^{\text{PPI}} u \right) du.
$$

Apply dominant convergence theorem to $\lim_{n\to\infty} \phi_{U_t}(s)$, we have:

$$
\begin{aligned}
&\lim_{n\to\infty} \phi_{U_t}(s) = \lim_{n\to\infty} \frac{1}{(2\pi)^{\frac{d}{2}} \det |V_{t-1}^{\text{PPI}}|} \int \exp\Big(is^T \mu(U_{t-1}) - \tfrac{1}{2} s^T \Sigma(U_{t-1}; r) s - \tfrac{1}{2} u^T V_{t-1}^{\text{PPI}} u \Big) du\\
&= \lim_{n\to\infty} \frac{1}{(2\pi)^{\frac{d}{2}} \det |V_{t-1}^{\text{PPI}}|} \int \exp\Big(is^T \sqrt{n}\big(G(\tfrac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1})\big)\\
&\quad - \tfrac{1}{2} s^T \Sigma_{\lambda_t, \frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}}\big(G(\tfrac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}); r\big) s - \tfrac{1}{2} u^T V_{t-1}^{\text{PPI}} u \Big) du\\
&= \frac{1}{(2\pi)^{\frac{d}{2}} \det |V_{t-1}^{\text{PPI}}|} \int \exp\Big(is^T \sqrt{n}\big(G(\tfrac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1})\big)
\end{aligned}
$$

$$
\begin{aligned}
&\quad - \tfrac{1}{2} s^T \Sigma_{\lambda_t, \frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}} \big(G(\tfrac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}); r\big) s - \tfrac{1}{2} u^T V_{t-1}^{\text{PPI}} u \Big) \, du \\
&= \frac{1}{(2\pi)^{\frac{d}{2}} \det |V_{t-1}^{\text{PPI}}|} \int \exp\Big(i s^T \nabla G(\theta_{t-1}) u - \tfrac{1}{2} s^T \Sigma_{\lambda_t, \theta_{t-1}} (G(\theta_{t-1}); r) s - \tfrac{1}{2} u^T V_{t-1}^{\text{PPI}} u \Big) \, du \\
&= \exp\Big(-\tfrac{1}{2} s^T \nabla G(\theta_{t-1}) V_{t+1}^{\text{PPI}} \nabla G(\theta_{t-1})^T s - \tfrac{1}{2} s^T \Sigma_{\lambda_t, \theta_{t-1}} (\theta_t; r) s \Big),
\end{aligned}
$$

which is the characteristic function of $\mathcal{N}(0, V_t^{\text{PPI}})$, where $V_t^{\text{PPI}} = \nabla G(\theta_{t-1}) V_{t-1}^{\text{PPI}} \nabla G(\theta_{t-1})^\top + \Sigma_{\lambda_t, \theta_{t-1}}(\theta_t; r)$. Here we use the fact that $\lim_{n \to \infty} \sqrt{n} \left(G(\frac{y}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1}) \right) = \nabla G(\theta_{t-1}) y$, and the dominant convergence theorem holds as we have

$$
\begin{aligned}
&| \exp(i s^T \sqrt{n} (G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}) - G(\theta_{t-1})) - \frac{1}{2} s^T \Sigma_{\lambda_t, \frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}} (G(\frac{U_{t-1}}{\sqrt{n}} + \theta_{t-1}); r) s - \frac{1}{2} u^T V_{t-1}^{\text{PPI}} u)| \\
&\leq | \exp(-\frac{1}{2} u^T V_{t-1}^{\text{PPI}} u)|.
\end{aligned}
$$

Thus we conclude by induction that

$$
\begin{aligned}
&U_t \xrightarrow{D} \mathcal{N}\left(0, V_t^{\text{PPI}}\right), \\
&V_t^{\text{PPI}} = \sum_{i=1}^{t} \left[\prod_{k=i}^{t-1} \nabla G(\theta_k) \right] \Sigma_{\lambda_{i-1}, \theta_{i-1}} (\theta_i; r) \left[\prod_{k=i}^{t-1} \nabla G(\theta_k) \right]^\top .
\end{aligned}
$$

And we have:

$$
\begin{aligned}
\nabla G(\theta_k) &= - \left[\mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}(\theta_k)} \nabla_\theta^2 \ell(x, y; \theta_{k+1}) \right]^{-1} \big(\nabla_{\tilde{\theta}} \mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}(\theta_k)} \nabla_\theta \ell(x, y; \theta_{k+1}) \big) \\
&= -H_{\theta_k}(\theta_{k+1})^{-1} \big(\nabla_{\tilde{\theta}} \mathop{\mathbb{E}}_{(x,y) \sim \mathcal{D}(\theta_k)} \nabla_\theta \ell(x, y; \theta_{k+1}) \big) \\
&= -H_{\theta_k}(\theta_{k+1})^{-1} \mathbb{E}_{z \sim \mathcal{D}(\theta_k)} [\nabla_\theta \ell(z, \theta_{k+1}) \nabla_\theta \log p(z, \theta_k)^\top].
\end{aligned}
$$

□

B Experimental Details

B.1 Additional Experimental Details

As described in Section 5, we construct simulation studies on a performative linear regression problem, where data are sampled from $D(\theta)$ as

$$y = \alpha^\top x + \mu^\top \theta + \nu,\ x \sim \mathcal{N}(\mu_x, \Sigma_x),\ \nu \sim \mathcal{N}(0, \sigma_y^2).$$

At each time step t, the label y_t is updated with $\hat{\theta}_{t-1}$ via the above equation, and then $\hat{\theta}_t$ is obtained by empirical repeated risk minimization with the updated data $z_t = (x_t, y_t)$. The objective of this task is to provide inference on an unbiased $\hat{\theta}_t$ with low variance, that is, the ground-truth θ_t is covered by the confidence region of $\hat{\theta}_t$ with high probability, and the width of this confidence region is small.

Given a set of labeled data, we can obtain the underlying θ_t as

$$\theta_t = (\Sigma_x + \mu_x\mu_x^\top + \gamma I_d)^{-1} \left(\mu_x\mu^\top\theta_{t-1} + (\Sigma_x + \mu_x\mu_x^\top)\alpha\right). \tag{4}$$

To compute the coverage and width of a confidence region $\hat{\mathcal{R}}_t(n, \delta)$ for θ_t, we run 1000 independent trials. For each trial j, we sample $\hat{\theta}_{t,j}$ together with with its estimated variance $\hat{V}_{t,j}$, and construct two-sided normal intervals for each coordinate $i = 1, \dots, d$:

$$\left[\hat{\theta}_{t,j}^{(i)} \pm q_{1-\frac{\delta}{2d}} \sqrt{\hat{V}_{t,j}^{(i)}/n}\right], \quad q_{1-\frac{\delta}{2d}} = \Phi^{-1}(1 - \frac{\delta}{2d}),$$

where $d = 2$ is the parameter dimension, $\delta = 0.1$ the significance level, n the data size, and Φ^{-1} the standard normal quantile. The interval width of each trial is averaged over d coordinate intervals, and we count this trial as covered if the ground-truth θ_t lies inside *all* d coordinate intervals simultaneously. Finally, we report the average width and coverage rate over all trials.

Similarly, to compute the coverage for performative stable point θ_{PS}, we can obtain the close-form θ_{PS} for this task as follows:

$$\theta_{\mathrm{PS}} = (\Sigma_x + \mu_x\mu_x^\top - \mu_x\mu^\top + \gamma I_d)^{-1}(\Sigma_x + \mu_x\mu_x^\top). \tag{5}$$

As defined in Corollary 3.6, the confidence region for θ_{PS} is constructed with

$$\hat{\mathcal{R}}_t(n, \delta) + \mathcal{B}\Big(0, 2B\big(\frac{\varepsilon\beta}{\gamma}\big)^t\Big),$$

where $\varepsilon = \|\mu\|_2$, $B = \|\theta_0 - \theta_{\mathrm{PS}}\|_2$, and

$$\beta = \max\left\{\max_{x\in\mathcal{X}}\{\|x\|_2^2 + \gamma\}, \max_{(x,y)\in(\mathcal{X},\mathcal{Y}),\theta\in\Theta}\{\sqrt{(x^\top\theta - y + \|x\|_2\|\theta\|_2)^2 + \|x\|_2^2}\}\right\}.$$

Here we take $\mathcal{X} = \{x : \|x\|_2^2 \leq 20\}$. Note that the closed form expressions for the update and the performatively stable point in Eq. 4 and Eq. 5 hold for any distribution of x with mean μ_x and variance Σ_x, and ν with mean 0 and variance σ_y^2. For easier calculation for the smoothness parameter, we truncate the normal distribution of (x, y) such that $\|x\|_2^2 \leq 20$. The mean and variance of the resulting truncated distribution can be well approximated by those of the original normal distribution due to the concentration of Gaussian.

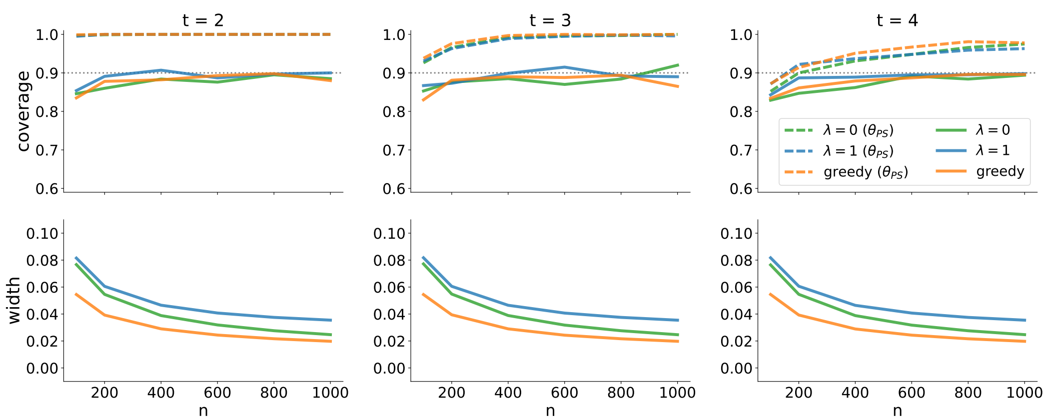
We run our experiments on NVIDIA GPUs A100 in a single-GPU setup.

B.2 Additional Experimental Results

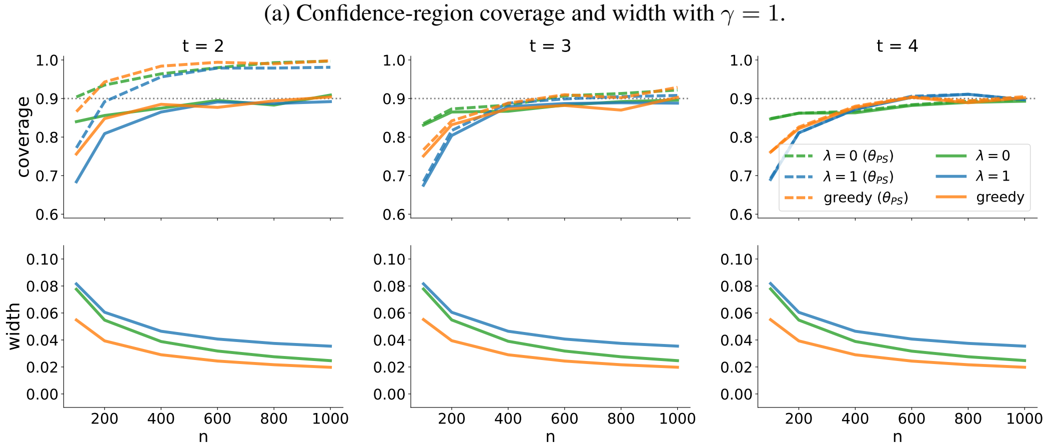
Ablation study on effects of γ. In Figure A1, we compare confidence-region performance under regularization strengths $\gamma = 1$ and $\gamma = 3$. Together with results of $\gamma = 2$ in Figure 1, we can find that as γ increases, the gap between the coverage for θ_t (solid curve) and the bias-adjusted coverage for θ_{PS} (dashed curve) vanishes more quickly across iterations. For example, at $t = 3$, the dashed and solid curves are tightly closed for $\gamma = 3$, while a substantial gap remains for $\gamma = 1$. This phenomenon derives that the larger γ yields a more strongly convex loss, which both accelerates convergence of the estimate $\hat{\theta}_t$ to its stable point and reduces the performative bias $\|\theta_t - \theta_{\mathrm{PS}}\|$. Consequently, the bias-awared intervals converge for θ_{PS} to the original ones for θ_t in fewer iterations when γ is larger.

Ablation study on effects of ε. In Figure A2, we compare confidence-region performance under sensitivity $\varepsilon \approx 0.003$ and $\varepsilon \approx 0.03$. We can find that as ε increases, the gap between the coverage for θ_t (solid curve) and the bias-adjusted coverage for θ_{PS} (dashed curve) vanishes more slowly across iterations. For example, for $\varepsilon \approx 0.003$, the dashed curves tightly upper-bound the solid curves at $t = 3$, whereas for $\varepsilon \approx 0.03$, a noticeable gap persists even at $t = 5$. This behavior is because a higher ε amplifies the performative shift (the dependence of the label distribution on θ), which increases the performative bias. That is, stronger sensitivity requires more iterations for $\hat{\theta}_t$ to approach its stable point, slowing down convergence of the two confidence regions.

Ablation study on effects of σ_y^2. In Figure A3, we compare confidence-region performance under noise level $\sigma_y^2 = 0.1$ and $\sigma_y^2 = 0.4$. We observe that across all settings, PPI with our greedy-selected $\hat{\lambda}$ is essentially never worse than either baseline $\lambda = 1$ or $\lambda = 0$. When the noise is low ($\sigma_y^2 = 0.1$), greedy $\hat{\lambda}$ behaves similarly to $\lambda = 0$, placing almost all weight on the true labels. Conversely, when the noise is high ($\sigma_y^2 = 0.4$), greedy $\hat{\lambda}$ behaves like $\lambda = 1$, relying more heavily on pseudo-labels to reduce variance. For the intermediate noise level $\sigma_y^2 = 0.2$ in Figure 1, greedy $\hat{\lambda}$ significantly outperforms both baselines by hitting the optimal bias–variance balance.

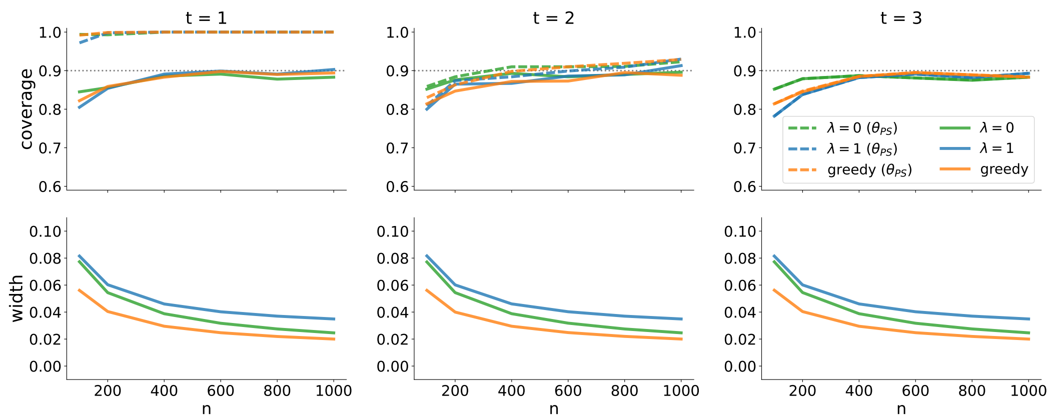


(a) Confidence-region coverage and width with $\gamma = 1$.

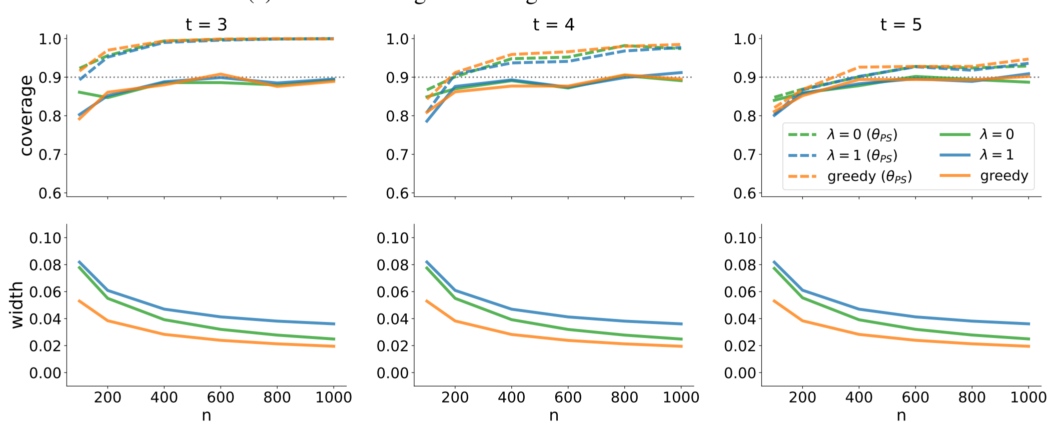


(b) Confidence-region coverage and width with $\gamma = 3$.

Figure A1: Confidence-region coverage (top row) and width (bottom row) with different choices of λ. The setup is the same as in Figure 1, only we change $\gamma = 1$ or $\gamma = 3$.

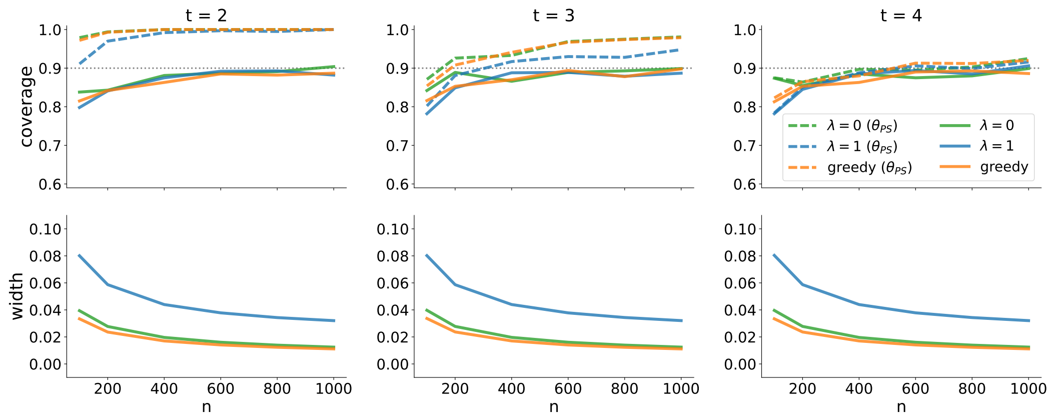


(a) Confidence-region coverage and width with $\varepsilon \approx 0.003$.

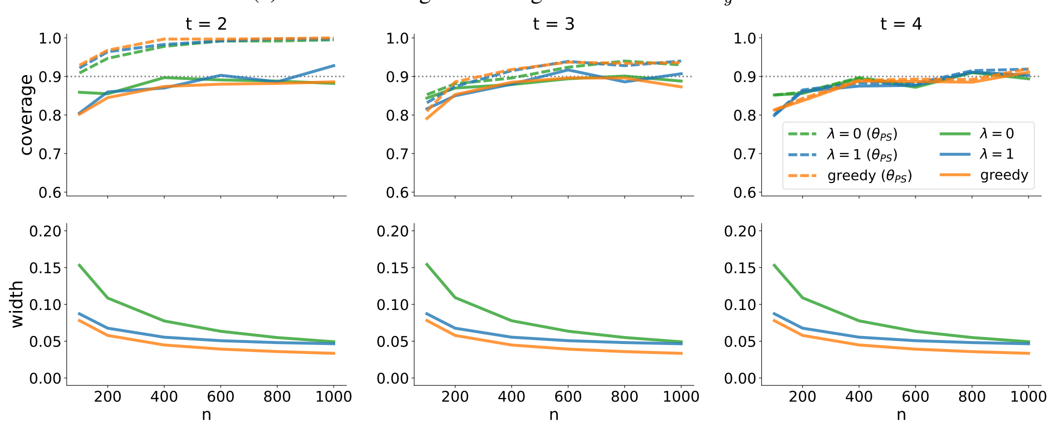


(b) Confidence-region coverage and width with $\varepsilon \approx 0.03$.

Figure A2: Confidence-region coverage (top row) and width (bottom row) with different choices of λ. The setup is the same as in Figure 1, only we change $\varepsilon \approx 0.003$ or $\varepsilon \approx 0.03$.



(a) Confidence-region coverage and width with $\sigma_y^2 = 0.1$.



(b) Confidence-region coverage and width with $\sigma_y^2 = 0.4$.

Figure A3: Confidence-region coverage (top row) and width (bottom row) with different choices of λ. The setup is the same as in Figure 1, only we change $\sigma_y^2 = 0.1$ or $\sigma_y^2 = 0.4$.