1–79

Statistical Inference After Adaptive Sampling in Non-Markovian Environments

Kelly W. Zhang KELLYWZHANG@SEAS.HARVARD.EDU
Department of Computer Science, Harvard University

Lucas Janson LJANSON@FAS.HARVARD.EDU
Department of Statistics, Harvard University

Susan A. Murphy SAMURPHY@FAS.HARVARD.EDU
Departments of Statistics and Computer Science, Harvard University

Abstract

There is a great desire to use adaptive sampling methods, such as reinforcement learning (RL) and bandit algorithms, for the real-time personalization of interventions in digital applications like mobile health and education. A major obstacle preventing more widespread use of such algorithms in practice is the lack of assurance that the resulting adaptively collected data can be used to reliably answer inferential questions, including questions about time-varying causal effects. Current methods for statistical inference on such data are insufficient because they (a) make strong assumptions regarding the environment dynamics, e.g., assume a contextual bandit or Markovian environment, or (b) require data to be collected with one adaptive sampling algorithm per user, which excludes data collected by algorithms that learn to select actions by pooling the data of multiple users. In this work, we make initial progress by introducing the *adaptive sandwich estimator* to quantify uncertainty; this estimator (a) is valid even when user rewards and contexts are non-stationary and highly dependent over time, and (b) accommodates settings in which an online adaptive sampling algorithm learns using the data of all users. Furthermore, our inference method is robust to misspecification of the reward models used by the adaptive sampling algorithm. This work is motivated by our work designing experiments in which RL algorithms are used to select actions, yet reliable statistical inference is essential for conducting primary analyses after the trial is over.

Keywords: adaptively collected data, reinforcement learning, bandits, micro-randomized trials, time-varying, causal inference, Z-estimation

1. Introduction

When designing a trial for a digital intervention in which reinforcement learning (RL) or other adaptive sampling algorithms are used to select actions, generally there are two primary considerations. The first is ensuring treatment interventions personalize and provide good user experiences, e.g., this could mean sending messages to users at opportune times. Mathematically this means minimizing regret or choosing the best actions with respect to some oracle policy. RL algorithms are designed specifically to optimize this objective. The second consideration is being able to utilize the user data collected by the adaptive sampling algorithm to perform statistical inference after the trial is over, e.g., construct a confidence interval for a treatment effect. Information gained from statistical inference is crucial for making decisions about whether to roll out or how to improve a given digital intervention. There are several aspects that make the above two considerations particularly challenging in the digital intervention context:

(a) **Complex Environment Dynamics:** A user's rewards and contexts can be non-stationary and highly dependent over time. The effects of actions also can be delayed, e.g., interventions can affect not only the immediate reward, but also affect a user's responsiveness in the future.

(b) **Low Signal Environments:** Rewards are noisy and intervention effects are generally small relative to the the noise variance.

Existing methods for statistical inference are insufficient for adaptively sampled data collected in such environments. Specifically, regarding (a), many inference methods for adaptively sampled data make strong assumptions regarding the environment dynamics, e.g., that the environment is a contextual bandit or Markovian (Hadad et al., 2021; Zhang et al., 2021; Bibaut et al., 2021b,a). Methods with such restrictive assumptions are less useful in digital intervention problems with more complex environment dynamics. This is especially true in the context of trials where the primary analysis, i.e., the foremost piece of reproducible knowledge gained by running a trial, should not have its validity hinge on strong environmental assumptions (such as a Markovian assumption) (Robins, 1986, 1997). However, existing statistical inference methods that address challenge (a), are not applicable to data collected by a large class of adaptive sampling algorithms designed to learn effectively in low signal environments, challenge (b). Specifically, these existing inference methods require independent user data trajectories (Boruvka et al., 2018; Qian et al., 2019), which excludes data collected by adaptive sampling algorithms that can potentially learn faster by pooling the data of multiple users. This is because adaptive sampling algorithms that learn across users induce dependence between the collected user data trajectories, since one user's reward affects how the algorithm updates and selects actions for other users at the next time-step.

Our Contribution In this work, we provide a novel inferential method for adaptively collected data that makes progress towards addressing the above challenges. Specifically we provide a method for constructing valid confidence regions with the following properties:

1. **Applicable to Non-Markovian Environments:** Our inference methods are valid even when user rewards and contexts are non-stationary and highly dependent over time. They are applicable to both Markovian and non-Markovian environments.

2. **Applicable to Datasets Collected by Algorithms that Learn Across Users:** Adaptive sampling algorithms that learn using the data of multiple users can potentially learn faster, but induce dependence between the collected user data trajectories. Our inference method is applicable to datasets collected by such algorithms because it accounts for induced dependence.

3. **Algorithm Agnostic:** We assume the adaptive sampling algorithm uses policies in a parametric policy class that is sufficiently smooth in the policy parameter and explores sufficiently. Besides these conditions, the validity of our inference method is not affected by potential misspecification of the adaptive sampling algorithm, e.g., if the RL algorithm incorrectly assumes a linear model for the reward or mistakenly assumes that the environment is Markovian, the validity of our inference method is not affected.

Specifically, we provide theory for inference via Z-estimators, which encompass most classical statistical estimators and can be used for estimating time-varying causal effects. We derive the asymptotic distribution of these estimators as the number of user data trajectories grows and construct confidence regions for parameters of interest using asymptotic approximations. We prove

that the standard sandwich estimator for the variance (Huber, 1967; Zeileis, 2006), which is valid if user data trajectories are independent, can underestimate the true variance when data is adaptively collected via algorithms that learn across users. We introduce the *adaptive sandwich estimator*, a corrected sandwich estimator that leads to consistent variance estimates under adaptive sampling. Besides accounting for the dependence between user data trajectories, another significant technical challenge is in accounting for how the estimation error in the policy parameters at one time-step impacts the error in future time-steps. A key tool we use to address this challenge is importance weights. These weights are purely a proof technique, and are *not* used to compute the estimator or to estimate the variance. Rather, we use these weights to implicitly define the policy parameter estimators as a function of the parameters of the policies used to select actions in previous time-steps. Finally, we illustrate our method's performance empirically via simulations.

1.1. Set-Up

We consider an adaptively collected batch dataset with T time-steps and n users. For each time-step $t \in [1\colon T]$ and user $i \in [1\colon n]$, we have random variables representing the vector-valued context $X_t^{(i)}$, the action $A_t^{(i)} \in \mathcal{A}$ for $|\mathcal{A}| < \infty$, and the reward $R_t^{(i)} \in \mathbb{R}$. Also, for each user we have a random variable representing the history where $\mathcal{H}_0^{(i)} \triangleq \emptyset$ and $\mathcal{H}_t^{(i)} := \{X_s^{(i)}, A_s^{(i)}, R_s^{(i)}\}_{s=1}^t$ for $t \in [1\colon T]$. We let $\mathcal{H}_t^{(1:n)} \triangleq \{\mathcal{H}_t^{(i)}\}_{i=1}^n$ represent the collective history for all users. We use the notation $X_{1:t}^{(i)} \triangleq \{X_s^{(i)}\}_{s=1}^t$ and $X_t^{(1:n)} \triangleq \{X_t^{(i)}\}_{i=1}^n$ to denote collections of random variables.

We use potential outcomes (Imbens and Rubin, 2015) to represent our counter-factual outcomes. We allow the potential outcomes for the rewards $R_t^{(i)}$ to depend on all actions taken on user i up to time-step t, $A_{1:t}^{(i)}$. This means $R_t^{(i)}$ has $\mathcal{A}^t$ different potential outcomes, $\{R_t^{(i)}(a_{1:t}) : a_{1:t} \in \mathcal{A}^t\}$. Similarly, contexts have potential outcomes $\{X_t^{(i)}(a_{1:t-1}) : a_{1:t-1} \in \mathcal{A}^{t-1}\}$. The observed variables are $R_t^{(i)} \triangleq R_t^{(i)}(A_{1:t}^{(i)})$ and $X_t^{(i)} \triangleq X_t^{(i)}(A_{1:t-1}^{(i)})$. We assume that potential outcomes are drawn independently across users $i \in [1\colon n]$ from an unknown distribution $\mathcal{P}$, i.e.,

$$\left\{X_t^{(i)}(a_{1:t-1}), R_t^{(i)}(a_{1:t}) : a_{1:t} \in \mathcal{A}^t\right\}_{t=1}^T \overset{i.i.d.}{\sim} \mathcal{P};\ \text{i.i.d over users } i \in [1\colon n]. \tag{1}$$

Note that our potential outcomes assumption encompasses both Markovian and non-Markovian environments, as it allows for a user's contexts and rewards to be non-stationary and dependent over time. Also, note our potential outcomes allow the context $X_t^{(i)}$ to contain all previous rewards $R_{1:t-1}^{(i)}$ and contexts $X_{1:t-1}^{(i)}$ from the same user. This type of potential outcome assumption is widely used in the longitudinal data analysis literature (Robins, 1986, 1997; Fitzmaurice et al., 2012).

For our statistical analyses we consider asymptotics as the number of users, n, goes to infinity and keep the total number of time-steps, T, fixed. This decision is motivated by our work in digital interventions, which is primarily concerned with using inference methods to draw scientific conclusions regarding a large population of individuals over a fixed period of time, e.g., a 90-day physical activity mobile health intervention for individuals with stage-1 hypertension (Liao et al., 2020).

We now provide the assumptions on the adaptive selection of the actions, that is, how the batch data was collected. For $t = 1$, we assume that there is a pre-specified policy π_1, where $\mathbb{P}(A_1^{(i)}|X_1^{(i)}) \triangleq \pi_1(A_1^{(i)}, X_1^{(i)})$. For each $t > 1$, the adaptive sampling algorithm can use all the observed data so far across users, $\mathcal{H}_{t-1}^{(1:n)}$, to form a policy $\hat{\pi}_t$. The policy $\hat{\pi}_t$ uses each user's current

context, $X_t^{(i)}$, to form action selection probabilities as follows for $a \in \mathcal{A}$,

$$\mathbb{P}\left(A_t^{(i)} = a \middle| X_t^{(1:n)}, \mathcal{H}_{t-1}^{(1:n)}\right) = \mathbb{P}\left(A_t^{(i)} = a \middle| X_t^{(i)}, \mathcal{H}_{t-1}^{(1:n)}\right) \triangleq \hat{\pi}_t\big(a, X_t^{(i)}\big). \tag{2}$$

We assume that $A_t^{(1:n)}$ are selected conditionally independently given policy $\hat{\pi}_t$ and contexts $X_t^{(1:n)}$. *Note that despite the i.i.d. potential outcomes assumption from Equation (1), the $\mathcal{H}_t^{(i)}$ are not independent over $i \in [1\colon n]$ because actions $A_t^{(i)}$ are selected using $\hat{\pi}_t$, which is formed using $\mathcal{H}_{t-1}^{(1:n)}$.* Additionally, we assume that policies $\hat{\pi}_t$ belong to a parametric class $\left\{\pi_t(\,\cdot\,; \beta_{t-1}) : \beta_{t-1} \in \mathbb{R}^{d_{t-1}}\right\}$, where $\pi_t(a, x; \beta_{t-1})$ is a probability of selecting an action a conditional on x. In particular, $\hat{\pi}_t$, is of the form $\pi_t(\,\cdot\,; \hat{\beta}_{t-1}^{(n)})$, where $\hat{\beta}_{t-1}^{(n)}$ is a function of all users' data prior to time t, $\mathcal{H}_{t-1}^{(1:n)}$. We will assume conditions under which $\hat{\beta}_{t-1}^{(n)}$ converges to a deterministic β_{t-1}^* as $n \to \infty$, and hence we call $\pi_t(\,\cdot\,; \beta_{t-1}^*)$, which we abbreviate as π_t^*, the *target policy* at time t.

1.2. Statistical Inference Objective

In this work we use the adaptively collected data described in Section 1.1 to construct a confidence region for a parameter, θ^*, in a model for the data. We assume that θ^* can be defined using a class of vector-valued functions $\psi\big(\mathcal{H}_T^{(i)}; \theta\big)$. In particular, θ^* satisfies

$$0 = \mathbb{E}_{\pi_{2:T}^*}\left[\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big)\right]. \tag{3}$$

Similarly, our estimator $\hat{\theta}^{(n)}$ satisfies $0 = \frac{1}{n}\sum_{i=1}^n \psi\big(\mathcal{H}_T^{(i)}; \hat{\theta}^{(n)}\big)$. This setup encompasses many types of standard estimators (e.g., least squares and maximum likelihood) and includes minimizers of differentiable loss functions.

The simplest example of θ^* is the *value* of the target policy, i.e., $\theta^* \triangleq \mathbb{E}_{\pi_{2:T}^*}\left[\frac{1}{T}\sum_{t=1}^T R_t^{(i)}\big(A_{1:t}^{(i)}\big)\right]$; this comes about by choosing $\psi\big(\mathcal{H}_T^{(i)}; \theta\big) \triangleq \frac{1}{T}\sum_{t=1}^T R_t^{(i)} - \theta$. The running example we use in this paper is a least-squares estimator in a binary action setting, $\mathcal{A} = \{0, 1\}$, with the following ψ:

$$\psi\big(\mathcal{H}_T^{(i)}; \theta\big) \triangleq \sum_{t=1}^T \left(R_t^{(i)} - \theta_0^\top X_t^{(i)} - A_t^{(i)}\theta_1^\top X_t^{(i)}\right) \begin{bmatrix} X_t^{(i)} \\ A_t^{(i)} X_t^{(i)} \end{bmatrix}. \tag{4}$$

Above, $\theta = [\theta_0, \theta_1]$ and the first entry of $X_t^{(i)}$ is 1 for all t, i. We are interested in constructing confidence regions for θ^*. We first characterize the asymptotic distribution of $\hat{\theta}^{(n)}$ as the number of users $n \to \infty$ and use that distribution to approximate the finite-sample distribution of $\hat{\theta}^{(n)}$.

Robust to Misspecification of the Statistical Model: Often in Z-estimation, ψ corresponds to the estimating equation (e.g., the score equation) for a parameter in a particular (possibly semi- or non-parametric) model for the data, and we can think of ψ as "correctly specified" if that model holds in our data. For our least squares example, ψ is correctly specified if $\mathbb{E}\big[R_t^{(i)} \big| X_t^{(i)}, A_t^{(i)}, \mathcal{H}_{t-1}^{(i)}\big] = \theta_0^{*,\top} X_t^{(i)} + A_t^{(i)}\theta_1^{*,\top} X_t^{(i)}$ w.p. 1 for all t. In the correctly specified case, θ^* does not depend on the target policies $\pi_{2:T}^*$. As is standard for Z-estimators, if ψ is not correctly specified, then θ^* is the best *projected* solution, i.e., the root of Equation (3). The projection is with respect to the distribution in which target policies $\pi_{2:T}^*$ are used to select actions.

Excursion Effects are a Key Use Case: Excursion effects, which are used for the primary analysis in micro-randomized trials (Boruvka et al., 2018; Qian et al., 2021), are a key use case for our inference method. The primary analysis for these trials concerns treatment effects under the study's sampling protocol. An example excursion effect is the following excursion from the target policy at time t:

$$\mathbb{E}_{\pi^*_{2:t-1}}\left[R_t^{(i)}\big(A_{1:t-1}^{(i)}, a_t = 1\big) - R_t^{(i)}\big(A_{1:t-1}^{(i)}, a_t = 0\big)\right].$$

In the simplified setting in which the reward $R_t^{(i)}$ only depends on the most recent action $A_t^{(i)}$, the excursion effect simplifies to the standard treatment effect $\mathbb{E}\big[R_t^{(i)}(a_t = 1) - R_t^{(i)}(a_t = 0)\big]$. Thus, the excursion effect above is a generalization of the standard treatment effect to environments in which all actions taken so far, $A_{1:t}^{(i)}$, can affect the distribution of the reward $R_t^{(i)}$.

1.3. Policy Class and Parameters

As discussed in Section 1.1, we assume that the batch data was adaptively collected using the learning policies $\hat{\pi}_{2:T}$. Recall that $\hat{\pi}_{2:T}$ are based on the statistics $\hat{\beta}_{1:T-1}^{(n)}$. We assume each $\hat{\beta}_t^{(n)}$ satisfies $\frac{1}{n}\sum_{i=1}^n \phi_t(\mathcal{H}_t^{(i)}; \hat{\beta}_t^{(n)}) = 0$ for a vector valued function ϕ_t. We define the policy target parameter β_t^* as the value of $\beta_t \in \mathbb{R}^{d_t}$ such that $\mathbb{E}_{\pi^*_{2:t}}\big[\phi_t(\mathcal{H}_t^{(i)}; \beta_t)\big] = 0$. For example, ϕ_t might be:

$$\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t\big) \triangleq \sum_{s=1}^{t}\big(R_s^{(i)} - \beta_{t,0}^\top X_s^{(i)} - A_s^{(i)}\beta_{t,1}^\top X_s^{(i)}\big)\begin{bmatrix} X_s^{(i)} \\ A_s^{(i)} X_s^{(i)}\end{bmatrix}; \tag{5}$$

this results in a least squares estimator, $\hat{\beta}_t^{(n)} = [\hat{\beta}_{t,0}^{(n)}, \hat{\beta}_{t,1}^{(n)}]$. Note however that the ϕ_t used by the adaptive sampling algorithm need not have any relation to the ψ used to define the parameter, θ^* in the statistical inference. We now discuss the policy classes,

$$\Big\{\pi_t\big(\,\cdot\,; \beta_{t-1}\big) : \beta_{t-1} \in \mathbb{R}^{d_{t-1}}\Big\}$$

for each $t \in [2\colon T]$. Recall that the adaptive sampling policy is $\hat{\pi}_t\big(\,\cdot\,\big) \triangleq \pi_t\big(\,\cdot\,; \hat{\beta}_{t-1}^{(n)}\big)$ and the target policy is $\pi_t^*\big(\,\cdot\,\big) \triangleq \pi_t\big(\,\cdot\,; \beta_{t-1}^*\big)$. We now discuss two key conditions that we assume on these policy classes. Below, for each $t \in [1\colon T-1]$, $B_t \subset \mathbb{R}^{d_t}$ is a bounded open ball around β_t^*.

Condition 1 (Minimum Exploration) *For some $\pi_{\min} > 0$, for all $t \in [1\colon T]$,*

$$\min_{a\in\mathcal{A}} \hat{\pi}_t\big(a, X_t^{(i)}\big) \geq \pi_{\min} \text{ w.p. } 1 \quad \text{and} \quad \inf_{\beta_{t-1}\in B_{t-1}} \min_{a\in\mathcal{A}} \pi_t\big(a, X_t^{(i)}; \beta_{t-1}\big) \geq \pi_{\min} \text{ w.p. } 1.$$

Condition 2 (Locally Lipschitz Policy Function) *For all $t \in [2\colon T]$, there exists a function $m_t\big(X_t^{(i)}\big)$ such that (i) $\mathbb{E}_{\pi^*_{2:t}}\big[m_t\big(X_t^{(i)}\big)\big] < \infty$ and (ii) for all $a \in \mathcal{A}$ and for any $\beta_t \in B_t$,*

$$\Big|\pi_t\big(a, X_t^{(i)}; \beta_{t-1}\big) - \pi_t\big(a, X_t^{(i)}; \beta_{t-1}^*\big)\Big| \leq m_t\big(X_t^{(i)}\big)\big\|\beta_{t-1} - \beta_{t-1}^*\big\|.$$

The first condition above is that the adaptive sampling policy used to selection actions, as well as all policies in a neighborhood of the target policy, produce action selection probabilities that are strictly bounded above zero for all actions. Note this condition excludes all deterministic policies,

which means policies that maximize the expected reward in standard stochastic bandit and Markov decision process environments are excluded. However, in general, the fewer structural assumptions that are placed on the environment, the more need there is for reward-maximizing algorithms to continually explore. For example, in non-stationary and adversarial sequential decision-making problem settings it is common both theoretically and in practice to prevent RL algorithms from allowing the action selection probabilities to go to zero for any action (Bubeck et al., 2012; Lattimore and Szepesvári, 2020; Cesa-Bianchi and Lugosi, 2006; Chandak et al., 2020) in order to ensure the algorithm can detect changes in the reward distribution.

The second condition is a smoothness condition on the policy function classes, which excludes policies that are a discontinuous function of parameters β_{t-1}. Although such a condition may appear rather mild, note that the reward-maximizing policy in a stochastic bandit problem is a discontinuous function of the margin because of the argmax operation, e.g., in a two-armed bandit setting with $\beta^* \triangleq \mathbb{E}[R_t(1)] - \mathbb{E}[R_t(0)]$, the optimal policy is $\mathbb{P}(A_t = 1) = \mathbb{I}_{\beta^* > 0}$. Despite this, as we discuss below, there are standard reinforcement learning algorithms developed for more complex environments (e.g., non-stationary) which satisfy this smoothness condition.

We now give an example of an online stochastic mirror descent algorithm, based on those from Lattimore and Szepesvári (2020, pg 361) and Bubeck et al. (2012), whose policy class satisfies Conditions 1 and 2 above. Note that for online stochastic mirror descent algorithms, $\hat{\pi}_t$ is an updated version of $\hat{\pi}_{t-1}$, which itself is an updated version of $\hat{\pi}_{t-2}$, and so on. This means that parameters of the class π_t must include those of $\pi_{t-1}, \pi_{t-2}, \ldots, \pi_2$. We will use slightly non-standard notation to represent this, $\hat{\pi}_t(\,\cdot\,) = \pi_t(\,\cdot\,; \hat{\beta}^{(n)}_{1:t-1})$, where each $\hat{\beta}^{(n)}_{t-1} = [\hat{\beta}^{(n)}_{t-1,0}, \hat{\beta}^{(n)}_{t-1,1}]$ is estimated using the least squares criterion from Equation (5). Since we consider a binary action setting, to characterize a policy it is sufficient to define the probability that action 1 is selected in each context.

$$\begin{aligned}\hat{\pi}_t\big(1, X^{(i)}_t\big) &= \pi_t\big(1, X^{(i)}_t; \hat{\beta}^{(n)}_{1:t-1}\big) \\ &= \text{argmin}_{p \in [\pi_{\min}, 1-\pi_{\min}]} \Big\{ \eta_t \big(\hat{\beta}^{(n),\top}_{t-1,0} X^{(i)}_t + p\, \hat{\beta}^{(n),\top}_{t-1,1} X^{(i)}_t \big) + \big(\hat{\pi}_{t-1}\big(1, X^{(i)}_t\big) - p \big)^2 \Big\}. \quad (6)\end{aligned}$$

Above, $\eta_t > 0$ is a learning rate and $\pi_{\min} \in (0, 0.5]$ is the minimum exploration rate. Note that $\hat{\beta}^{(n),\top}_{t-1,0} X^{(i)}_t + p\, \hat{\beta}^{(n),\top}_{t-1,1} X^{(i)}_t$ is an estimate of the expectation of $R^{(i)}_t$ given $X^{(i)}_t, \mathcal{H}^{(i)}_{t-1}$ when $A^{(i)}_t$ is selected with probability p. The term $\big(\hat{\pi}_{t-1}(1, X^{(i)}_t) - p\big)^2$ is a Bregman divergence and can be replaced by other Bregman divergences, e.g., KL-divergence. By Equation (6) above, we can derive

$$\pi_t\big(1, X^{(i)}_t; \hat{\beta}^{(n)}_{1:t-1}\big) = \text{Clip}_{\pi_{\min}}\left(\pi_{t-1}\big(1, X^{(i)}_t; \hat{\beta}^{(n)}_{1:t-2}\big) - \frac{1}{2} \eta_t \hat{\beta}^{(n),\top}_{t-1,1} X^{(i)}_t \right), \quad (7)$$

where $\text{Clip}_{\pi_{\min}}(x) \triangleq \min((x, \pi_{\min}), 1 - \pi_{\min})$. We can also show that Condition 2 holds because $\big|\pi_t\big(1, X^{(i)}_t; \beta_{1:t-1}\big) - \pi_t\big(1, X^{(i)}_t; \beta^*_{1:t-1}\big)\big| \leq \frac{1}{2}\eta_t \big\|X^{(i)}_t\big\| \, \big\|\beta_{t-1,1} - \beta^*_{t-1,1}\big\|$ for any $\beta_{1:t-1} \in \mathbb{R}^{\sum_{s=1}^{t-1} d_s}$. See Lemma 3 for derivations of the above results.

2. Related Work

Recently, many novel inference methods have been developed for adaptively collected data focused on multi-armed and contextual bandit environments. These include inference methods via asymptotic approximations (Hadad et al., 2021; Zhang et al., 2021; Bibaut et al., 2021b,a; Zhang et al.,

2020; Zhan et al., 2021; Chen et al., 2020; Deshpande et al., 2018) as well as approaches that use high probability bounds (Howard et al., 2018; Karampatziakis et al., 2021; Brennan et al., 2020; Abbasi-Yadkori et al., 2011). These works for the most part consider asymptotics as $T \to \infty$. These methods are more restrictive in that they assume an underlying contextual bandit environment that does not allow a user's potential outcomes to be dependent over time. However, these methods are more general than ours in that they put fewer restrictions on the adaptive sampling policies used to collect the batch data, e.g., many allow the action selection probabilities to go to zero for some actions and do not require the policy class to be smooth in its parameter.

Another area of related work are methods for inferring excursion effects (Boruvka et al., 2018; Qian et al., 2019). These methods assume the same underlying potential outcomes model, Equation (1), which allows for non-stationarity and dependent outcomes over time. However, they also assume the batch data was collected using separate adaptive sampling algorithms for each user. Our work can be considered an extension of these works to the setting in which the batch data was collected by a single adaptive sampling algorithm that learns using the data from multiple users.

We utilize techniques from the classical literature on empirical processes (Van der Vaart, 2000; Van Der Vaart and Wellner, 1996). In particular, as we will discuss in Section 3.2, we derive a maximal inequality for weighted empirical processes on the adaptively collected data described in Section 1. Recent work Bibaut et al. (2021b) develop a novel maximal inequality for adaptively collected data in the contextual bandit environment. Besides the differences in the underlying environment assumptions, our maximal inequality results also differ from theirs because they consider asymptotics as $T \to \infty$, while we let T be fixed and consider asymptotics as $n \to \infty$.

3. Asymptotic Results

Note that if the batch data were collected using the fixed target policies $\pi^*_{2:T}$, rather than the adaptive policies $\hat{\pi}_{2:T}$, then the data trajectories would be independent across users, i.e., $\mathcal{H}_T^{(i)}$ would be i.i.d. across $i \in [1\colon n]$. In that i.i.d. setting, we could use standard asymptotic normality results for Z-estimators (Van der Vaart, 2000, Theorem 5.21) to get that $\hat{\theta}^{(n)}$ is asymptotically normal with the standard sandwich variance, i.e., $\sqrt{n}\big(\hat{\theta}^{(n)} - \theta^*\big) \xrightarrow{D} \mathcal{N}\big(0, \dot{\Psi}^{-1} M (\dot{\Psi}^{-1})^\top\big)$ with "bread" $\dot{\Psi} \triangleq \mathbb{E}_{\pi^*_{2:T}}\big[\frac{\partial}{\partial \theta^*}\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big)\big]$ and "meat" $M \triangleq \mathbb{E}_{\pi^*_{2:T}}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*_T\big)^{\otimes 2}\big]$. Note $x^{\otimes 2} \triangleq xx^\top$. However, in our adaptively collected data setting where $\hat{\pi}_{2:T}$ are used to select actions, we show that the limiting variance is different, specifically,

$$\sqrt{n}\left(\hat{\theta}^{(n)} - \theta^*\right) \xrightarrow{D} \mathcal{N}\left(0, \dot{\Psi}^{-1} M^{\text{adaptive}} (\dot{\Psi}^{-1})^\top\right), \tag{8}$$

where $M^{\text{adaptive}} \triangleq \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big) + \dot{\Psi} \sum_{t=1}^{T-1} \big(\frac{\partial \theta^*}{\partial \beta_t^*}\big) \dot{\Phi}_t^{-1} \phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\right\}^{\otimes 2}\right]$ and $\dot{\Phi}_t \triangleq \mathbb{E}_{\pi^*_{2:t}}\big[\frac{\partial}{\partial \beta_t^*}\phi\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\big]$. We call the limiting variance in Equation (8), the *adaptive* sandwich variance. Comparing M^{adaptive} and M, we can interpret the term $\dot{\Psi} \sum_{t=1}^{T-1} \big(\frac{\partial \theta^*}{\partial \beta_t^*}\big) \dot{\Phi}_t^{-1} \phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)$ as the "cost" or "inflation" in variance due to using the estimated $\hat{\pi}_{2:T}$ to select actions rather than $\pi^*_{2:T}$. See Section 4 for simulation results demonstrating the performance of our approach.

A key technique we use in this work is implicitly defining functions, which allow us to define derivative terms like $\frac{\partial \theta^*}{\partial \beta_t^*}$ in M^{adaptive}. Recall β_2^* is defined as $\mathbb{E}_{\pi_2^*}\big[\phi_2\big(\mathcal{H}_2^{(i)}; \beta_2^*\big)\big] = 0$. If we allow π_2^* to vary in this expression by varying its inputs β_1, then β_2^* can be considered a

function of β_1, where $0 = \mathbb{E}_{\pi_2(\beta_1)}\big[\phi_2\big(\mathcal{H}_2^{(i)}; \beta_2^*(\beta_1)\big)\big] = 0$. We overload notation by using β_2^* to refer to the vector $\beta_2^*(\beta_1^*)$. Similarly, we can consider $\beta_3^*(\,\cdot\,)$ to be a function of $\beta_{1:2}$ where $0 = \mathbb{E}_{\pi_2(\beta_1),\pi_3(\beta_2)}\big[\phi_3\big(\mathcal{H}_2^{(i)}; \beta_3^*(\beta_{1:2})\big)\big] = 0$; we also use β_3^* to refer to the vector $\beta_3^*(\beta_{1:2}^*)$. Continuing this process we get that θ^* is an implicitly defined function of $\beta_{1:T-1}$ such that

$$0 = \mathbb{E}_{\pi_2(\beta_1),\pi_3(\beta_2),\ldots,\pi_{T-1}(\beta_{T-1})}\Big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big(\beta_1, \beta_2, \ldots, \beta_{T-1}\big)\big)\Big] \tag{9}$$

and θ^* is $\theta^*(\beta_{1:T-1}^*)$. This allows us to define $\frac{\partial \theta^*}{\partial \beta_t^*}$ using implicit differentiation. See Lemma 4 for sufficient conditions for the derivative terms in M^{adaptive} to exist.

Additionally, we define $\hat{\theta}^{(n)}$ as an implicit function of $\beta_{1:T-1}$, analogously to how we defined θ^* as an implicit function of $\beta_{1:T-1}$ above. The key tool we use to do this is importance weighting. Interestingly, these weights are purely a proof technique, and are *not* necessary for computing the estimator or for estimating the variance of the estimator. Define for any $\beta_t, \beta_t' \in \mathbb{R}^{d_t}$,

$$W_{t+1}^{(i)}(\beta_t, \beta_t') \triangleq \frac{\pi_{t+1}(A_{t+1}^{(i)}, X_{t+1}^{(i)}; \beta_t)}{\pi_{t+1}(A_2^{(i)}, X_{t+1}^{(i)}; \beta_t')}. \tag{10}$$

Specifically, $\hat{\beta}_2^{(n)}(\beta_1)$ solves $0 = \frac{1}{n}\sum_{i=1}^n W_2^{(i)}\big(\beta_1, \hat{\beta}_1^{(n)}\big)\phi_2\big(\mathcal{H}_2^{(i)}; \hat{\beta}_2^{(n)}(\beta_1)\big)$. $\hat{\beta}_3^{(n)}(\beta_{1:2})$ solves $0 = \frac{1}{n}\sum_{i=1}^n W_2^{(i)}\big(\beta_1, \hat{\beta}_1^{(n)}\big)W_3^{(i)}\big(\beta_2, \hat{\beta}_2^{(n)}\big)\phi_3\big(\mathcal{H}_3^{(i)}; \hat{\beta}_3^{(n)}(\beta_{1:2})\big)$, and $\hat{\theta}^{(n)}(\beta_{1:t-1})$ solves $0 = \frac{1}{n}\sum_{i=1}^n \big\{\prod_{t=1}^{T-1} W_{t+1}^{(i)}\big(\beta_t, \hat{\beta}_t^{(n)}\big)\big\}\psi\big(\mathcal{H}_T^{(i)}; \hat{\theta}^{(n)}(\beta_{1:t-1})\big)$; $\hat{\beta}_t^{(n)} = \hat{\beta}_t^{(n)}(\hat{\beta}_{1:t-1}^{(n)})$ and $\hat{\theta}^{(n)} = \hat{\theta}^{(n)}(\hat{\beta}_{1:T-1}^{(n)})$.

3.1. Formal Results and Conditions

We first state our asymptotic results formally. Then we introduce and discuss the conditions we use. Throughout, B_1 is some bounded open ball around β_1^*; B_2 is a bounded open ball that contains all $\beta_2^*(\beta_1)$ for all $\beta_1 \in B_1$; and for each $t > 2$, B_t is a bounded open ball that contains $\beta_t^*(\beta_{1:t-1})$ for all $\beta_{1:t-1} \in B_{1:t-1} \triangleq B_1 \times B_2 \times \ldots \times B_{t-1}$. Similarly, Θ is a bounded ball that contains $\theta^*(\beta_{1:T-1})$ for all $\beta_{1:T-1} \in B_{1:T-1}$. Throughout, for functions of $B_{1:t-1}$, e.g., $\beta_t^*(\,\cdot\,)$, we use the sup norm, $\|\beta_t^*(\,\cdot\,)\|_{B_{1:t-1}} \triangleq \sup_{\beta_{1:t-1} \in B_{1:t-1}} \|\beta_t^*(\beta_{1:t-1})\|_2$.

Theorem 1 (Consistency) *We consider the setting of Section 1. Under Conditions 1, 3, 4, 5, 6, and 7, $\hat{\theta}^{(n)} \xrightarrow{P} \theta^*$ and $\hat{\beta}_t^{(n)} \xrightarrow{P} \beta_t^*$ for all $t \in [1\colon T-1]$. Moreover, $\|\hat{\theta}^{(n)}(\,\cdot\,) - \theta^*(\,\cdot\,)\|_{B_{1:T-1}} \xrightarrow{P} 0$ and $\|\hat{\beta}_t^{(n)}(\,\cdot\,) - \beta_t^*(\,\cdot\,)\|_{B_{1:-1}} \xrightarrow{P} 0$ for all $t \in [1\colon T-1]$.*

Theorem 2 (Asymptotic Normality) *We consider the setting of Section 1. Assuming Conditions 1, 2, 3, 4, 5, and 8 and the conclusions of Theorem 1, Equation (8) holds.*

Conditions 1 and 2 were discussed above. Condition 3 below requires that slightly larger than the fourth moments of functions $\phi_t(\,\cdot\,; \beta_t^*)$ and $\psi(\,\cdot\,; \theta^*)$ are bounded.

Condition 3 (Finite Moments) *For some $\alpha > 0$, for all $t \in [1\colon T]$,*

$$\mathbb{E}_{\pi_{2:t}^*}\Big[\big\|\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\big\|_1^{4+\alpha}\Big] < \infty \quad \text{and} \quad \mathbb{E}_{\pi_{2:T}^*}\Big[\big\|\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big)\big\|_1^{4+\alpha}\Big] < \infty.$$

Our next condition concerns the functions $f_T(\,\cdot\,; \beta_{1:T-1}, \theta) \triangleq \big(\prod_{t=2}^T \pi_t(\,\cdot\,; \beta_{t-1})\big)\psi\big(\,\cdot\,; \theta\big)$ and the functions $f_t(\,\cdot\,; \beta_{1:t}) \triangleq \big(\prod_{s=2}^{t-1} \pi_s(\,\cdot\,; \beta_{s-1})\big)\phi_t\big(\,\cdot\,; \beta_t\big)$.

Condition 4 (Lipschitz Estimating Functions) *There exists a function* $g_T\big(\mathcal{H}_T^{(i)}\big)$ *such that (i)* $\mathbb{E}_{\pi^*_{2:T}}\big[g_T\big(\mathcal{H}_T^{(i)}\big)^2\big] < \infty$ *and (ii) for all* $\beta_{1:T-1}, \beta'_{1:T-1} \in B_{1:T-1}$*, and all* $\theta, \theta' \in \Theta$*,*

$$\Big\| f_T\big(\mathcal{H}_T^{(i)}; \beta_{1:T-1}, \theta\big) - f_T\big(\mathcal{H}_T^{(i)}; \beta'_{1:T-1}, \theta'\big) \Big\| \leq g_T\big(\mathcal{H}_T^{(i)}\big) \left\| \big[\beta_{1:T-1}, \theta\big] - \big[\beta'_{1:T-1}, \theta'\big] \right\|.$$

Also, for all $t \in [2\colon T]$*, there exists a function* $g_t\big(\mathcal{H}_t^{(i)}\big)$ *such that (i)* $\mathbb{E}_{\pi^*_{2:t}}\big[g_t\big(\mathcal{H}_t^{(i)}\big)^2\big] < \infty$ *and (ii) for all* $\beta_{1:t}, \beta'_{1:t} \in B_{1:t}$*,* $\|f_t(\mathcal{H}_t^{(i)}; \beta_{1:t}, \theta) - f_t(\mathcal{H}_t^{(i)}; \beta_{1:t})\| \leq g_t(\mathcal{H}_t^{(i)})\|\beta_{1:t} - \beta'_{1:t}\|$*.*

One key use of Condition 4 is to control the bracketing complexity of the function class $\mathcal{F}_{T,c_T} \triangleq \big\{c_T^\top f_T(\,\cdot\,; \beta_{1:T-1}, \theta) : \beta_{1:T-1} \in B_{T-1}, \theta \in \Theta\big\}$ for any fixed $c_T \in \mathbb{R}^{d_T}$. By Example 19.7 of Van der Vaart (2000), Condition 4 implies that $\int_0^1 \sqrt{\log N_{[\,]}\big(\epsilon, \mathcal{F}_{T,c_T}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon < \infty$, where $\mathcal{P}_{\pi^*_{2:T}}$ is the distribution of potential outcomes $\mathcal{P}$ under policies $\pi^*_{2:T}$. The finite bracketing integral property is closely related Donsker conditions in the i.i.d. case (Van der Vaart, 2000, Theorem 19.5).

The next condition concerns the differentiability of $\beta_t^*(\,\cdot\,)$, $\theta^*(\,\cdot\,)$. Recall $\beta_2^*(\,\cdot\,)$ is a function of $\beta_1 \in B_1$ and $\beta_3^*(\,\cdot\,)$ is a function of $\beta_{1:2} \in B_{1:2}$. To represent the function of $\beta_1 \in B_1$ that outputs $\beta_3^*\big(\beta_1, \beta_2^*(\beta_1)\big)$ we define functions of the form $\beta_3^{*,[1]}$ where the superscript represents the number of β arguments the function takes. Specifically, we let $\beta_2^{*,[1]}(\beta_1) \triangleq \beta_2^*(\beta_1)$, $\beta_3^{*,[1]}(\beta_1) \triangleq \beta_3^*\big(\beta_1, \beta_2^{*,[1]}(\beta_1)\big)$, and $\beta_3^{*,[2]}(\beta_{1:2}) \triangleq \beta_3^*(\beta_{1:2})$. For general t, $\beta_t^{*,[s]}(\beta_{1:s}) \triangleq \beta_t^*\big(\beta_{1:s}, \beta_{s+1:t-1}^{*,[s]}(\beta_{1:s})\big)$ for any $s < t$. Following this pattern, we can also define $\theta^{*,[s]}(\,\cdot\,)$ for $s < T$ which takes arguments $\beta_{1:s} \in B_{1:s}$. We say $\theta^{*,[s-1]}(\,\cdot\,)$ is Fréchet differentiable with respect to $\beta_s^*(\,\cdot\,)$ for all $s < T$ if there exists a function $\frac{\partial \theta^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)} : B_{1:s-1} \mapsto \mathbb{R}^{d_T \times d_s}$ such that for any function $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto \mathbb{R}^{d_s}$,

$$\left\| \theta^{*,[s]}(\,\cdot\,, \beta_s(\,\cdot\,)) - \theta^{*,[s-1]}(\,\cdot\,) - \frac{\partial \theta^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}\big(\beta_s(\,\cdot\,) - \beta_s^*(\,\cdot\,)\big) \right\|_{B_{1:s-1}} = o\Big(\big\|\beta_s(\,\cdot\,) - \beta_s^*(\,\cdot\,)\big\|_{B_{1:s-1}}\Big).$$

Condition 5 (Fréchet Differentiability) $\theta^{*,[s-1]}(\,\cdot\,)$ *is Fréchet differentiable with respect to* $\beta_s^*(\,\cdot\,)$ *and continuous in* $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ *for all* $s < T$*, and for all* $t \in [1\colon T-1]$*,* $\beta_t^{*,[s-1]}(\,\cdot\,)$ *is Fréchet differentiable with respect to* $\beta_s^*(\,\cdot\,)$ *and continuous in* $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ *for all* $s < t$*. Also, the derivative functions* $\frac{\partial \theta^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ *and* $\frac{\partial \beta_t^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ *are continuous in their arguments* $\beta_{1:s-1} \in B_{1:s-1}$*.* $\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ *is Fréchet differentiable with respect to* $\theta^*(\,\cdot\,)$*, and* $\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ *is Fréchet differentiable with respect to* $\beta_t^*(\,\cdot\,)$ *for all* $t \in [1\colon T-1]$*. Also, derivative functions* $\frac{\partial}{\partial \theta^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ *and* $\frac{\partial}{\partial \beta_t^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ *are continuous in their arguments,* $\beta_{1:T-1} \in B_{1:T-1}$ *and* $\beta_{1:t-1} \in B_{1:t-1}$*, respectively.*

Fréchet differentiability is often used to prove asymptotic normality resultw, e.g., Van der Vaart (2000, Theorem 19.26). Theorem 1 is proved using modifications of standard techniques used in the i.i.d. setting (Van der Vaart, 2000, Theorem 5.9), and relies on the following two conditions. Condition 6 ensures that θ^* is a unique root. Assumptions similar to Condition 7 hold when ψ, ϕ_t are derivatives of convex criteria (Van Der Vaart and Wellner, 1996; Bura et al., 2018).

Condition 6 (Well-Separated Solution) *For any* $\epsilon > 0$*, there exists some* $\eta_\epsilon > 0$ *such that for all* $\beta_{1:T-1} \in B_{1:T-1}$*,*

$$\inf_{\theta\in\mathbb{R}^{d_T}:\|\theta-\theta^*(\beta_{1:T-1})\|>\epsilon} \left\|\mathbb{E}_{\pi_2(\beta_1),\pi_3(\beta_2),\ldots,\pi_T(\beta_{T-1})}\left[\psi(\mathcal{H}_T^{(i)};\theta)\right]\right\| > \eta_\epsilon.$$

Similarly, for $t \in [1\colon T-1]$, *for any* $\epsilon > 0$ *there exists some* $\eta_{t,\epsilon} > 0$ *such that for all* $\beta_{1:t-1} \in B_{1:t-1}$, $\inf_{\beta_t\in\mathbb{R}^{d_t}:\|\beta_t-\beta_t^*(\beta_{1:t-1})\|>\epsilon} \left\|\mathbb{E}_{\pi_2(\beta_1),\pi_3(\beta_2),\ldots,\pi_t(\beta_{t-1})}\left[\phi_t\left(\mathcal{H}_t^{(i)};\beta_t\right)\right]\right\| > \eta_{t,\epsilon}$.

Condition 7 (Compact Parameter Space) $\mathbb{P}\left(\left\{\hat{\theta}^{(n)}(\beta_{1:T-1}) : \beta_{1:T-1} \in B_{1:T-1}\right\} \subset \Theta\right) \to 1$ *and* $\mathbb{P}\left(\left\{\hat{\beta}_t^{(n)}(\beta_{1:t-1}) : \beta_{1:t-1} \in B_{1:t-1}\right\} \subset B_t\right) \to 1$ *for all* $t \in [1\colon T-1]$.

Our final condition is that the "bread" terms in the sandwich variance estimator are uniformly positive definite. Recall that $\dot{\Psi} \triangleq \frac{\partial}{\partial\theta^*}\mathbb{E}_{\pi^*_{2:T}}\left[\psi\left(\mathcal{H}_T^{(i)};\theta^*\right)\right]$ in Equation (8).

Condition 8 (Positive Definite Bread) $\frac{\partial}{\partial\theta^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\left[\psi\left(\mathcal{H}_T^{(i)};\theta^*(\,\cdot\,)\right)\right]$ *is finite and positive definite uniformly over* $\beta_{1:T-1} \in B_{1:T-1}$. *Also,* $\frac{\partial}{\partial\beta_t^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\left[\phi_t\left(\mathcal{H}_t^{(i)};\beta_t^*(\,\cdot\,)\right)\right]$ *is finite and positive definite uniformly over* $\beta_{1:t-1} \in B_{1:t-1}$ *for all* $t \in [1\colon T-1]$.

3.2. Proof Sketch of Asymptotic Normality Result

We focus on the $T = 2$ case for this proof sketch because it is illustrative of the main proof techniques we use. See Appendix C for the full proof for general T. The desired asymptotically normality result Equation (8) can be rewritten as follows when $T = 2$:

$$\sqrt{n}\left(\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \dot{\Psi}^{-1}M^{\text{adaptive}}(\dot{\Psi}^{-1})^\top\right). \tag{11}$$

Let $\Sigma_{1,1} \triangleq \mathbb{E}_{\pi_2^*}\left[\phi_1(\mathcal{H}_1^{(i)};\beta_1^*)^{\otimes 2}\right]$, $\Sigma_{2,2} \triangleq \mathbb{E}_{\pi_2^*}\left[\psi(\mathcal{H}_2^{(i)};\theta^*)^{\otimes}\right]$, $\Sigma_{1,2} \triangleq \mathbb{E}_{\pi_2^*}\left[\phi_1(\mathcal{H}_1^{(i)};\beta_1^*)\psi(\mathcal{H}_2^{(i)};\theta^*)^\top\right]$, and $\Sigma_{2,1} = \Sigma_{1,2}^\top$. To prove Equation (11), we use the result:

$$\sqrt{n}\begin{bmatrix}\hat{\beta}_1^{(n)} - \beta_1^* \\ \hat{\theta}^{(n)}(\hat{\beta}_1^{(n)}) - \theta^*(\hat{\beta}_1^{(n)})\end{bmatrix} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix}\dot{\Phi}_1^{-1}\Sigma_{1,1}(\dot{\Phi}_1^{-1})^\top & \dot{\Phi}_1^{-1}\Sigma_{1,2}(\dot{\Psi}^{-1})^\top \\ \dot{\Psi}^{-1}\Sigma_{2,1}(\dot{\Phi}_1^{-1})^\top & \dot{\Psi}^{-1}\Sigma_{2,2}(\dot{\Psi}^{-1})^\top\end{bmatrix}\right). \tag{12}$$

Here we show why (11) follows from (12). Recall the function $\theta^*(\,\cdot\,) : \mathbb{R}^{d_1} \mapsto \mathbb{R}^{d_2}$. By Condition 5, $\frac{\partial\theta^*}{\partial\beta_1^*} = \frac{\partial}{\partial\beta_1^*}\theta^*\left(\beta_1^*\right)$ exists. Thus, $\sqrt{n}\left(\theta^*\left(\hat{\beta}_1^{(n)}\right) - \theta^*(\beta_1^*)\right) = \frac{\partial\theta^*}{\partial\beta_1^*}\sqrt{n}\left(\hat{\beta}_1^{(n)} - \beta_1^*\right) + o_P(1)$ by the Delta method (Van der Vaart, 2000, Theorem 3.1). Thus, by Equation (12) and Slutsky's theorem,

$$\sqrt{n}\begin{bmatrix}\theta^*\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right) \\ \hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\hat{\beta}_1^{(n)}\right)\end{bmatrix} \xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix}\frac{\partial\theta^*}{\partial\beta_1^*}\dot{\Phi}_1^{-1}\Sigma_{1,1}(\dot{\Phi}_1^{-1})^\top\frac{\partial\theta^*}{\partial\beta_1^*}^\top & \frac{\partial\theta^*}{\partial\beta_1^*}\dot{\Phi}_1^{-1}\Sigma_{1,2}(\dot{\Psi}^{-1})^\top \\ \dot{\Psi}^{-1}\Sigma_{2,1}(\dot{\Phi}_1^{-1})^\top\frac{\partial\theta^*}{\partial\beta_1^*}^\top & \dot{\Psi}^{-1}\Sigma_{2,2}(\dot{\Psi}^{-1})^\top\end{bmatrix}\right). \tag{13}$$

Note that we can decompose the difference $\sqrt{n}\left[\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right)\right]$ as follows:

$$\sqrt{n}\left[\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right)\right] = \sqrt{n}\Bigg[\underbrace{\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\hat{\beta}_1^{(n)}\right)}_{\hat{\theta}^{(n)}\text{ vs. }\theta^*}\Bigg] + \sqrt{n}\Bigg[\underbrace{\theta^*\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right)}_{\hat{\beta}_1^{(n)}\text{ vs. }\beta_1^*}\Bigg].$$

Furthermore, the differences on the right hand side above are equivalent to those on the left hand side of Equation (13). Thus, by Equation (13), we have the following result:

$$\sqrt{n}\left(\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) - \theta^*\left(\beta_1^*\right)\right) \xrightarrow{D} \mathcal{N}\left(0, V\right),$$

where $V \triangleq \frac{\partial \theta^*}{\partial \beta_1^*}\dot{\Phi}_1^{-1}\Sigma_{1,1}(\dot{\Phi}_1^{-1})^\top \frac{\partial \theta^*}{\partial \beta_1^*}^\top + \frac{\partial \theta^*}{\partial \beta_1^*}\dot{\Phi}_1^{-1}\Sigma_{1,2}(\dot{\Psi}^{-1})^\top + \dot{\Psi}^{-1}\Sigma_{2,1}(\dot{\Phi}_1^{-1})^\top \frac{\partial \theta^*}{\partial \beta_1^*}^\top + \dot{\Psi}^{-1}\Sigma_{2,2}(\dot{\Psi}^{-1})^\top$. Note by rearranging terms we can show that $V = \dot{\Psi}^{-1}M^{\text{adaptive}}(\dot{\Psi}^{-1})^\top$, where recall that $M^{\text{adaptive}} \triangleq \mathbb{E}_{\pi_2^*}\left[\left\{\psi(\mathcal{H}_2^{(i)};\theta^*) + \dot{\Psi}\left(\frac{\partial \theta^*}{\partial \beta_1^*}\right)\dot{\Phi}_1^{-1}\phi_1(\mathcal{H}_1^{(i)};\beta_1^*)\right\}^{\otimes 2}\right]$. Thus, it remains only to show Equation (12).

Showing Equation (12) Holds: Our proof to show Equation (12) has a structure similar to that of classical Z-estimator asymptotic normality proofs for i.i.d. data (Van der Vaart, 2000, Theorem 5.21). However, since the user data trajectories $\mathcal{H}_T^{(i)}$ are not independent, our proof differs in key ways, which we highlight below. We first define the following functions of $\beta_1 \in \mathbb{R}^{d_1}$ and $\theta \in \mathbb{R}^{d_2}$:

$$\Psi(\beta_1,\theta) \triangleq \mathbb{E}\left[W_2^{(i)}(\beta_1,\hat{\beta}_1^{(n)})\psi(\mathcal{H}_2^{(i)};\theta)\right] \quad \text{and} \quad \hat{\Psi}^{(n)}(\beta_1,\theta) \triangleq \frac{1}{n}\sum_{i=1}^{n} W_2^{(i)}(\beta_1,\hat{\beta}_1^{(n)})\psi(\mathcal{H}_2^{(i)};\theta).$$

Recall we use $\mathbb{E}$ to refer to expectations with respect to the distribution of the observed data, which means $\hat{\pi}_2$ are used to select actions. Thus, $\mathbb{E}\left[W_2^{(i)}(\beta_1,\hat{\beta}_1^{(n)})\psi(\mathcal{H}_2^{(i)};\theta)\right] = \mathbb{E}_{\pi_2(\beta_1)}\left[\psi(\mathcal{H}_2^{(i)};\theta)\right]$. The most crucial result we use to show Equation (12) is the result of Lemma 16:

$$\sqrt{n}c_2^\top\left[\hat{\Psi}^{(n)}\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right) - \Psi\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right)\right]$$
$$= \sqrt{n}c_2^\top\left[\hat{\Psi}^{(n)}\left(\beta_1^*,\theta^*\left(\beta_1^*\right)\right) - \Psi\left(\beta_1^*,\theta^*\left(\beta_1^*\right)\right)\right] + o_P(1), \quad (14)$$

where $c_2 \in \mathbb{R}^{d_2}$ is any fixed vector. By Theorem 1 we have that $\hat{\beta}_1^{(n)} \xrightarrow{P} \beta_1^*$ and $\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right) \xrightarrow{P} \theta^*\left(\beta_1^*\right)$. Thus, intuitively, Equation (14) is saying that the random function $\sqrt{n}c_2^\top\left[\hat{\Psi}^{(n)}(\,\cdot\,) - \Psi(\,\cdot\,)\right]$ can have its arguments $\left[\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right]$ replaced with their limits, $\left[\beta_1^*,\theta^*\left(\beta_1^*\right)\right]$, without affecting this random function's asymptotic distribution. The proof of Equation (14) features a maximal inequality for the stochastic process: $\left\{\sqrt{n}c_2^\top\left[\hat{\Psi}^{(n)}(\beta_1,\theta) - \Psi(\beta_1,\theta)\right] : \beta_1 \in B_1, \theta \in \Theta\right\}$. Note by our definition of the class $\mathcal{F}_{2,c_2}$ (see text below Condition 4), the above stochastic process is equivalent to $\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\hat{\pi}_2(A_2^{(i)},X_2^{(i)})^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\hat{\pi}_2(A_2^{(i)},X_2^{(i)})^{-1}f(\mathcal{H}_2^{(i)})\right]\right) : f \in \mathcal{F}_{2,c_2}\right\}$. We use Condition 4, which restricts the bracketing complexity of $\mathcal{F}_{2,c_2}$, to prove a maximal inequality for the previous stochastic process. Note our maximal inequality differs from classical ones because our observations $\mathcal{H}_T^{(i)}$ are dependent over $i \in [1\colon n]$. See Lemma 11 for details.

The left hand side of Equation (14) can be simplified as follows:

$$\sqrt{n}c_2^\top\left[\hat{\Psi}^{(n)}\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right) - \Psi\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right)\right]$$
$$= -\sqrt{n}c_2^\top\left[\Psi\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right) - \Psi\left(\hat{\beta}_1^{(n)},\theta^*\left(\hat{\beta}_1^{(n)}\right)\right)\right]$$

Above, the equality holds since $\hat{\Psi}^{(n)}\left(\hat{\beta}_1^{(n)},\hat{\theta}^{(n)}\left(\hat{\beta}_1^{(n)}\right)\right) = 0$ and $\Psi\left(\hat{\beta}_1^{(n)},\theta^*\left(\hat{\beta}_1^{(n)}\right)\right) = 0$ by the definitions of $\hat{\theta}^{(n)}(\,\cdot\,)$ and $\theta^*(\,\cdot\,)$ respectively. The equality below holds because by Condition 5, $\Psi(\,\cdot\,,\theta_2^*(\,\cdot\,))$ is Fréchet differentiable with respect to $\theta_2^*(\,\cdot\,)$ (using the sup norm over B_1) and $\mathbb{P}(\hat{\beta}_1^{(n)} \in B_1) \to 1$ by Condition 7.

$$= -c_2^\top\frac{\partial}{\partial\theta_2^*(\hat{\beta}_1^{(n)})}\Psi\left(\hat{\beta}_1^{(n)},\theta_2^*(\hat{\beta}_1^{(n)})\right)\sqrt{n}\left(\hat{\theta}^{(n)}(\hat{\beta}_1^{(n)}) - \theta^*\left(\hat{\beta}_1^{(n)}\right)\right)$$
$$+ \sqrt{n}o_P\left(\left\|\hat{\theta}^{(n)}(\,\cdot\,) - \theta^*(\,\cdot\,)\right\|_{B_1}\right) + o_P(1).$$

We can show that $\sqrt{n}o_P\big(\big\|\hat{\theta}^{(n)}(\hat{\beta}_1^{(n)}) - \theta^*\big(\hat{\beta}_1^{(n)}\big)\big\|\big) = o_P(1)$ using Lemma 8. By Condition 5 the derivative $\frac{\partial}{\partial \theta_2^*(\beta_1)}\Psi\big(\beta_1, \theta_2^*(\beta_1)\big)$ is continuous in $\beta_1 \in B_1$. Since $\hat{\beta}_1^{(n)} \xrightarrow{P} \beta_1^*$, by the continuous mapping theorem, $\frac{\partial}{\partial \theta_2^*(\hat{\beta}_1^{(n)})}\Psi\big(\hat{\beta}_1^{(n)}, \theta_2^*(\hat{\beta}_1^{(n)})\big) \xrightarrow{P} \frac{\partial}{\partial \theta_2^*(\beta_1^*)}\Psi\left(\beta_1^*, \theta_2^*(\beta_1^*)\right) = \dot{\Psi}$.

$$= -c_2^\top \dot{\Psi}\sqrt{n}\left(\hat{\theta}^{(n)}(\hat{\beta}_1^{(n)}) - \theta^*\big(\hat{\beta}_1^{(n)}\big)\right) + o_P(1). \tag{15}$$

Recall the goal is to prove Equation (12). To do this, consider

$$c_1^\top \dot{\Phi}_1\sqrt{n}\big(\hat{\beta}_1^{(n)} - \beta_1^*\big) + c_2^\top \dot{\Psi}\sqrt{n}\left(\hat{\theta}^{(n)}(\hat{\beta}_1^{(n)}) - \theta^*\big(\hat{\beta}_1^{(n)}\big)\right)$$
$$= -\sqrt{n}c_1^\top \hat{\Phi}_1^{(n)}\left(\beta_1^*\right) - \sqrt{n}c_2^\top \hat{\Psi}^{(n)}\left(\beta_1^*, \theta^*\big(\beta_1^*\big)\right) + o_P(1), \tag{16}$$

where $\hat{\Phi}_1^{(n)}(\beta_1) \triangleq \frac{1}{n}\sum_{i=1}^n \phi_1(\mathcal{H}_1^{(i)}; \beta_1^*)$. The above holds since $\dot{\Phi}_1\sqrt{n}\big(\hat{\beta}_1^{(n)} - \beta_1^*\big) = -\sqrt{n}\hat{\Phi}_1^{(n)}(\beta_1^*) + o_P(1)$ by Theorem 19.5 of Van der Vaart (2000), and by Equation (15) which we showed equals the left hand side of Equation (14), combined with $\Psi\left(\beta_1^*, \theta^*\big(\beta_1^*\big)\right) = 0$ by definition of $\theta^*(\,\cdot\,)$. Finally, Equation (12) holds by Equation (16), Slutsky's Theorem, the Cramer–Wold device, and the invertibility of $\dot{\Phi}_1$ and $\dot{\Psi}$ by Condition 8.

4. Simulation Results

We compare the empirical coverage of confidence regions constructed using both the standard sandwich and adaptive sandwich variance estimators. We consider a binary action setting in which all previous actions $A_{1:t-1}^{(i)}$ can affect of the mean of $R_t^{(i)}$, however, these delayed effects are decaying, i.e., more recent actions have a larger effects. We set $T = 50$. See Appendix A.1 for more details. As seen in Table 1, the adaptive sandwich estimator consistently outperforms the standard sandwich variance estimator. Moreover, the performance gap increases with the magnitude of the delayed effects (κ larger means larger delayed effects). Although we see some undercoverage for the adaptive estimator with large delayed effects when $n = 100$, this is shown to be only finite-sample behavior as the coverage is very close to 95% when $n = 1000$; when the delayed effects are small, the $n = 100$ coverage is already at 95%. Note larger delayed effects means previous actions have a larger effect on the current time-step's reward distribution, so we expect larger delayed effects to increase the magnitude of $\frac{\partial \theta^*}{\partial \beta_t^*}$ (if terms $\frac{\partial \theta^*}{\partial \beta_t^*}$ are zero then the two limiting variances are equivalent).

Table 1: **Empirical Coverage of Confidence** 95% **Regions for** θ^***.** All standard errors < 0.01.

	$n = 100, \kappa = 1$	$n = 1000, \kappa = 1$	$n = 100, \kappa = 0.5$	$n = 1000, \kappa = 0.5$
Sandwich	68.16%	71.24%	93.6%	93.24%
Adaptive Sandwich	88.72%	94.28%	95.44%	95.76%

5. Discussion

The greatest limitation of this work is that it does not apply to adaptively collected batch datasets in which the policies used to collect the data are (i) not smooth in their parameters or (ii) allow the amount of exploration to go to zero. As discussed in Section 1.3, many common RL algorithms for

bandit and Markov decision process settings do not satisfy our smoothness and exploration conditions. However, the studies in digital interventions motivating this work will use adaptive sampling algorithms that do satisfy these conditions. Needed work includes the derivation of efficient estimators based on adaptively collected data and methods for forming valid confidence intervals for the value or for an excursion effect under a different policy unrelated to that used to collect the data.

Acknowledgments

Research reported in this paper was supported by NIH grants numbers P50DA05403, P41EB028242, and UG3DE028723. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This material is based upon work supported by the National Science Foundation grant number NSF CBET–2112085 and by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE1745303. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

Aurélien Bibaut, Antoine Chambaz, Maria Dimakopoulou, Nathan Kallus, and Mark van der Laan. Post-contextual-bandit inference. *NeurIPS*, 2021a.

Aurélien Bibaut, Antoine Chambaz, Maria Dimakopoulou, Nathan Kallus, and Mark van der Laan. Risk minimization from adaptively collected data: Guarantees for supervised and policy learning. *NeurIPS*, 2021b.

Audrey Boruvka, Daniel Almirall, Katie Witkiewitz, and Susan A Murphy. Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association*, 113 (523):1112–1121, 2018.

Jennifer Brennan, Ramya Korlakai Vinayak, and Kevin Jamieson. Estimating the number and effect sizes of non-null hypotheses. In *International Conference on Machine Learning*, pages 1123–1133. PMLR, 2020.

Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41–1. JMLR Workshop and Conference Proceedings, 2012.

Efstathia Bura, Sabrina Duarte, Liliana Forzani, Ezequiel Smucler, and Mariela Sued. Asymptotic theory for maximum likelihood estimates in reduced-rank multivariate generalized linear models. *Statistics*, 52(5):1005–1024, 2018.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Yash Chandak, Georgios Theocharous, Shiv Shankar, Martha White, Sridhar Mahadevan, and Philip Thomas. Optimizing for the future in non-stationary mdps. In *International Conference on Machine Learning*, pages 1414–1425. PMLR, 2020.

Haoyu Chen, Wenbin Lu, and Rui Song. Statistical inference for online decision making: In a contextual bandit setting. *Journal of the American Statistical Association*, pages 1–16, 2020.

Yash Deshpande, Lester Mackey, Vasilis Syrgkanis, and Matt Taddy. Accurate inference for adaptive linear models. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1194–1203, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Aryeh Dvoretzky. Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.

Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

Vitor Hadad, David A. Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the National Academy of Sciences*, 118(15), 2021. ISSN 0027-8424. doi: 10.1073/pnas.2014602118. URL https://www.pnas.org/content/118/15/e2014602118.

Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Uniform, nonparametric, non-asymptotic confidence sequences. *arXiv preprint arXiv:1810.08240*, 2018.

Peter J Huber. Under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Weather modification*, volume 5, page 221. Univ of California Press, 1967.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Nikos Karampatziakis, Paul Mineiro, and Aaditya Ramdas. Off-policy confidence sequences. *arXiv preprint arXiv:2102.09540*, 2021.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Peng Liao, Kristjan Greenewald, Predrag Klasnja, and Susan Murphy. Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22, 2020.

Tianchen Qian, Hyesun Yoo, Predrag Klasnja, Daniel Almirall, and Susan A Murphy. Estimating time-varying causal excursion effect in mobile health with binary outcomes. *arXiv preprint arXiv:1906.00528*, 2019.

Tianchen Qian, Ashley E Walton, Linda M Collins, Predrag Klasnja, Stephanie T Lanza, Inbal Nahum-Shani, Mashifiqui Rabbi, Michael A Russell, Maureen A Walton, Hyesun Yoo, et al. The micro-randomized trial for developing digital interventions: Experimental design and data analysis considerations. *arXiv preprint arXiv:2107.03544*, 2021.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7 (9-12):1393–1512, 1986.

James M Robins. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*, pages 69–117. Springer, 1997.

Aad W Van der Vaart. *Asymptotic Statistics*, volume 3. Cambridge University Press, 2000.

Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer, 1996.

Achim Zeileis. Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(1):1–16, 2006.

Ruohan Zhan, Vitor Hadad, David A Hirshberg, and Susan Athey. Off-policy evaluation via adaptive weighting with data from contextual bandits. *arXiv preprint arXiv:2106.02029*, 2021.

Kelly W Zhang, Lucas Janson, and Susan A Murphy. Inference for batched bandits. *NeurIPS 2020*, 2020.

Kelly W Zhang, Lucas Janson, and Susan A Murphy. Statistical inference with m-estimators on adaptively collected datas. *NeurIPS 2021*, 2021.

Contents

Appendix A. Simulations and Examples

A.1. Simulation Results

We consider the binary action setting in which $\hat{\theta}^{(n)}$ and $\hat{\beta}_t^{(n)} = [\hat{\beta}_{t,0}^{(n)}, \hat{\beta}_{t,1}^{(n)}]$ are both least squares estimators as defined in Equations (4) and (5). We use policy classes of the form $\pi_t(1, X_t^{(i)}; \beta_{t-1}) = \text{Clip}_{0.1}\big(\text{expit}(\beta_{t-1,1}^\top X_t^{(i)})\big)$ where $\text{Clip}_{0.1}(x) \triangleq \min\big(\max(x, 0.1), 1 - 0.1\big)$. The context is the previous time-step's reward, i.e., $X_t^{(i)} = \big[1, R_{t-1}^{(i)}\big]$. Reward potential outcomes are generated as follows:

$$R_t^{(i)}\big(A_{1:t}^{(i)}\big) = \alpha_{t,0}^\top A_{1:t}^{(i)} + \alpha_{t,1}^\top A_{1:t}^{(i)} \cdot R_{t-1}^{(i)}\big(A_{1:t-1}^{(i)}\big) + \epsilon_t^{(i)}$$

where $A_{1:t}^{(i)} = \big[A_1^{(i)}, A_2^{(i)}, ..., A_t^{(i)}\big]$. For all t, i, $\epsilon_t^{(i)} \sim \mathcal{N}(0, 1)$ marginally, however, $\text{Corr}(\epsilon_t^{(i)}, \epsilon_t^{(i)}) = 0.5^{|t-s|/2}$, which means each user's reward errors are correlated over time. The above means that all previous actions $A_{1:t-1}^{(i)}$ can affect of the mean of $R_t^{(i)}$, however, we will have delayed effects are decaying, i.e., more recent actions have a larger effects. Specifically we let $\alpha_{t,0} \triangleq \kappa\alpha_t$ and $\alpha_{t,1} \triangleq 0.5\kappa\alpha_t$ for $\alpha_t = [e^{-(t-2)}, ..., e^{-2}, e^{-1}, e^0, 0]$. We vary the magnitude of the delayed effects by setting $\kappa = 1$ (large delayed effects) and $\kappa = 0.5$ (small delayed effects). We set $T = 50$. We construct the adaptive sandwich variance estimator by using an empirical estimator for the adaptive sandwich variance $\dot{\Psi}^{-1} M^{\text{adaptive}} (\dot{\Psi}^{-1})^\top$ (see Lemma 4 for explicit derivations of the derivative terms in M^{adaptive}). Our simulation results are averaged over 2500 Monte Carlo repetitions.

A.2. Lemma 3: Stochastic Mirror Descent Example

Lemma 3 (Stochastic Mirror Descent Example) *We consider the stochastic mirror descent algorithm example from Section 1.3. We show that Equation (7) and that Condition 2 hold under the condition that* $\mathbb{E}_{\pi^*_{2:t}}\left[\left\|X_t^{(i)}\right\|\right] < \infty$ *for all* $t \in [1\colon T]$.

Proof of Lemma 3 Recall that the stochastic mirror descent algorithm described earlier in Section 1.3 has action selection probabilities of the following form (see Equation (6)):

$$\begin{aligned}\hat{\pi}_t\big(1, X_t^{(i)}\big) &= \pi_t\big(1, X_t^{(i)}; \hat{\beta}_{1:t-1}^{(n)}\big) \\ &= \text{argmin}_{p \in [\pi_{\min}, 1-\pi_{\min}]} \Big\{ \eta_t \big(\hat{\beta}_{t-1,0}^{(n),\top} X_t^{(i)} + p\, \hat{\beta}_{t-1,1}^{(n),\top} X_t^{(i)}\big) + \big(\hat{\pi}_{t-1}\big(1, X_t^{(i)}\big) - p\big)^2 \Big\},\end{aligned}$$

where above $\eta_t > 0$ is a learning rate and $\pi_{\min} \in (0, 0.5]$. Note that exploration Condition 1 is satisfied because the algorithm constrains action selection probabilities between $[\pi_{\min}, 1 - \pi_{\min}]$. By taking the derivative of the following criterion with respect to p:

$$\eta_t \big(\hat{\beta}_{t-1,0}^{(n),\top} X_t^{(i)} + p\, \hat{\beta}_{t-1,1}^{(n),\top} X_t^{(i)}\big) + \big(\hat{\pi}_{t-1}\big(1, X_t^{(i)}\big) - p\big)^2 \tag{17}$$

we have

$$\eta_t\, \hat{\beta}_{t-1,1}^{(n),\top} X_t^{(i)} - 2\hat{\pi}_{t-1}\big(1, X_t^{(i)}\big) + 2p.$$

Since second derivative of the criterion from Equation (17) with respect to p is $2 > 0$, the global minimizer of the criterion (not restricted to $[\pi_{\min}, 1-\pi_{\min}]$) is $p = \hat{\pi}_{t-1}\big(1, X_t^{(i)}\big) - \frac{1}{2}\eta_t\, \hat{\beta}_{t-1,1}^{(n),\top} X_t^{(i)}$. Also note that the criterion from Equation (17) is convex because its derivative is strictly increasing in p. Note that the constrained minimizer of a convex function either equals the global minimizer or

is on the boundary of the constraint space. Thus we have that the constrained minimizer, $\hat{\pi}_t\big(1, X_t^{(i)}\big)$, equals the following:

$$\pi_t\big(1, X_t^{(i)}; \hat{\beta}_{1:t-1}^{(n)}\big) = \text{Clip}_{\pi_{\min}}\left(\hat{\pi}_{1:t-1}(1, X_t^{(i)}) - \frac{1}{2}\eta_t \hat{\beta}_{t-1,1}^{(n),\top} X_t^{(i)}\right),$$

where $\text{Clip}_{\pi_{\min}}(x) \triangleq \min\big(\max(x, \pi_{\min}), 1-\pi_{\min}\big)$. Thus, we have shown that Equation (7) mentioned in Section 1.3 holds.

We now show that this algorithm class satisfies Condition 2. Note that for any $\beta_{1:t-1} \in \mathbb{R}^{\sum_{s=1}^{t-1} d_t}$,

$$\begin{aligned}
&\left|\pi_t\big(1, X_t^{(i)}; \beta_{1:t-1}\big) - \pi_t\big(1, X_t^{(i)}; \beta_{1:t-1}^*\big)\right| \\
&= \left|\text{Clip}_{\pi_{\min}}\left(\hat{\pi}_{t-1}(1, X_t^{(i)}) - \frac{1}{2}\eta_t \beta_{t-1,1}^{\top} X_t^{(i)}\right) - \text{Clip}_{\pi_{\min}}\left(\hat{\pi}_{t-1}(1, X_t^{(i)}) - \frac{1}{2}\eta_t \beta_{t-1,1}^{*,\top} X_t^{(i)}\right)\right|. \quad (18)
\end{aligned}$$

Note that for any real numbers x, y that $\left|\text{Clip}_{\pi_{\min}}(x) - \text{Clip}_{\pi_{\min}}(y)\right| \leq |x-y|$. This is because

- If $x, y \in [\pi_{\min}, 1-\pi_{\min}]$, then $\left|\text{Clip}_{\pi_{\min}}(x) - \text{Clip}_{\pi_{\min}}(y)\right| = |x-y|$.
- If $x, y < \pi_{\min}$ or $x, y > 1-\pi_{\min}$, then $0 = \left|\text{Clip}_{\pi_{\min}}(x) - \text{Clip}_{\pi_{\min}}(y)\right| \leq |x-y|$.
- If $x > \pi_{\min}$ and $y < \pi_{\min}$, then $\left|\text{Clip}_{\pi_{\min}}(x) - \text{Clip}_{\pi_{\min}}(y)\right| \leq x - \text{Clip}_{\pi_{\min}}(y) < x - y = |x-y|$.
- If $x < 1-\pi_{\min}$ and $y > 1-\pi_{\min}$, then $\left|\text{Clip}_{\pi_{\min}}(x) - \text{Clip}_{\pi_{\min}}(y)\right| \leq \text{Clip}_{\pi_{\min}}(y) - x < y - x = |x-y|$.

Thus, we have that Equation (18) can be upper bounded by the following:

$$\begin{aligned}
&\leq \left|\hat{\pi}_{t-1}(1, X_t^{(i)}) - \frac{1}{2}\eta_t \beta_{t-1,1}^{\top} X_t^{(i)} - \hat{\pi}_{t-1}(1, X_t^{(i)}) + \frac{1}{2}\eta_t \beta_{t-1,1}^{*,\top} X_t^{(i)}\right| \\
&= \left|\frac{1}{2}\eta_t \left(\beta_{t-1,1} - \beta_{t-1,1}^*\right)^{\top} X_t^{(i)}\right| \leq \frac{1}{2}\eta_t \left\|X_t^{(i)}\right\| \left\|\beta_{t-1,1} - \beta_{t-1,1}^*\right\|
\end{aligned}$$

The last inequality above holds by Cauchy-Schwartz. By the above, Condition 2 holds since $\mathbb{E}\left[\left\|X_t^{(i)}\right\|\right] < \infty$. ■

A.3. Lemma 4: Sufficient Conditions for $\frac{\partial \theta^*}{\partial \beta_t^*}$ to Exist

Lemma 4 (Sufficient Conditions for $\frac{\partial \theta^*}{\partial \beta_t^*}$ to Exist) *Under the following conditions, the implicit derivative $\frac{\partial \theta^*}{\partial \beta_t^*}$ exists for all $t \in [1\colon T-1]$:*

- $\dot{\Psi} \triangleq \mathbb{E}_{\pi_{2:T}^*}\left[\frac{\partial}{\partial \theta^*}\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big)\right]$ *and* $\dot{\Phi}_t \triangleq \mathbb{E}_{\pi_{2:t}^*}\left[\frac{\partial}{\partial \beta_t^*}\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\right]$ *for all* $t \in [1\colon T-1]$ *are finite and invertible.*
- *Let* $\dot{\pi}_t^*(A_t^{(i)}, X_t^{(i)}) \triangleq \frac{\partial}{\partial \beta_{t-1}^*}\pi_t\big(A_t^{(i)}, X_t^{(i)}; \beta_{t-1}^*\big) \in \mathbb{R}^{d_{t-1}}$. $\mathbb{E}_{\pi_{2:T}^*}\left[\psi(\mathcal{H}_T^{(i)}; \theta^*)\frac{\dot{\pi}_t^*(A_t^{(i)}, X_t^{(i)})^{\top}}{\pi_t^*(A_t^{(i)}, X_t^{(i)})}\right]$ *exists for all* $t \in [1\colon T-1]$, *and* $\mathbb{E}_{\pi_{2:t}^*}\left[\phi_t(\mathcal{H}_t^{(i)}; \beta_t^*)\frac{\dot{\pi}_s^*(A_s^{(i)}, X_s^{(i)})^{\top}}{\pi_s^*(A_s^{(i)}, X_s^{(i)})}\right]$ *exists for any* $s < t$.

Proof of Lemma 4 For notational convenience, we let $\theta_T \triangleq \theta$ and $\theta_t \triangleq \beta_t$ for all $t \in [1\colon T-1]$. This means that $\theta_T^* \triangleq \theta^*$ and $\theta_t^* \triangleq \beta_t^*$. Additionally, we let $\psi_t \triangleq \phi_t$ for all $t \in [1\colon T-1]$, so $\psi_t(\mathcal{H}_t^{(i)}; \theta_t) = \phi_t(\mathcal{H}_t^{(i)}; \beta_t)$.

Note that by $\frac{\partial \theta_3^*}{\partial \theta_1^*}$ we mean $\frac{\partial \theta_3^*}{\partial \theta_1^*} = \frac{\partial \theta_3^*(\theta_{1:2}^*)}{\partial \theta_1^*} = \frac{\partial \theta_3^*(\theta_1^*, \theta_2(\theta_1^*))}{\partial \theta_1^*}$, where we differentiate through the first argument of θ_3^* as well as $\theta_2(\theta_1^*)$, the second argument.

It is sufficient to show that $\frac{\partial \theta_T^*}{\partial \theta_t^*}$ exists for any $t < T$. Our proof approach is to use an induction argument. Specifically, we will show the base case that $\frac{\partial \theta_2^*}{\partial \theta_1^*}$ exists. Then we will show the inductive step that $\frac{\partial \theta_t^*}{\partial \theta_s^*}$ exists for all $s < t$ whenever $\frac{\partial \theta_{t-1}^*(\theta_{1:t-2}^*)}{\partial \theta_{s'}^*}$ exists for all $s' < t-1$.

Base Case: Recall that θ_2^* is defined as follows:

$$0 = \mathbb{E}_{\pi_2^*}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\right]$$

We use implicit differentiation to derive $\frac{\partial \theta_2^*}{\partial \theta_1}$. We will also use the following weights defined earlier in Equation (10): $W_2^{(i)}(\theta_1, \hat{\theta}_1^{(n)}) \triangleq \frac{\pi_2(A_2^{(i)}, X_2^{(i)}; \theta_1)}{\pi_2(A_2^{(i)}, X_2^{(i)}; \hat{\theta}_1^{(n)})}$.

$$0 = \frac{\partial}{\partial \theta_1}\mathbb{E}_{\pi_2^*}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\right] = \frac{\partial}{\partial \theta_1^*}\mathbb{E}\left[W_2^{(i)}(\theta_1^*, \hat{\theta}_1^{(n)})\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\right]$$

$$= \mathbb{E}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\left(\frac{\partial}{\partial \theta_1} W_2^{(i)}(\theta_1^*, \hat{\theta}_1^{(n)})\right)\right] + \mathbb{E}\left[W_2^{(i)}(\theta_1^*, \hat{\theta}_1^{(n)})\frac{\partial}{\partial \theta_1^*}\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\right]$$

$$= \mathbb{E}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*(\theta_1))\frac{\dot{\pi}_2^*(A_2^{(i)}, X_2^{(i)})^\top}{\pi_2(A_2^{(i)}, X_2^{(i)}; \hat{\theta}_1^{(n)})}\right] + \mathbb{E}\left[W_2^{(i)}(\theta_1, \hat{\theta}_1^{(n)})\frac{\partial}{\partial \theta_2^*(\theta_1)}\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*(\theta_1))\right]\frac{\partial \theta_2^*(\theta_1)}{\partial \theta_1}$$

$$= \mathbb{E}_{\pi_2^*}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\frac{\dot{\pi}_2^*(A_2^{(i)}, X_2^{(i)})^\top}{\pi_2^*(A_2^{(i)}, X_2^{(i)})}\right] + \underbrace{\mathbb{E}_{\pi_2^*}\left[\frac{\partial}{\partial \theta_2^*}\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\right]}_{=\dot{\Psi}_2}\frac{\partial \theta_2^*}{\partial \theta_1}$$

Note that by our assumptions, $\dot{\Psi}_2$ is finite and invertible, so by the above results,

$$\frac{\partial \theta_2^*}{\partial \theta_1^*} = -\dot{\Psi}_2^{-1}\mathbb{E}\left[\psi_2(\mathcal{H}_2^{(i)}; \theta_2^*)\frac{\dot{\pi}_2(A_2^{(i)}, X_2^{(i)}; \theta_1^*)^\top}{\pi_2^*(A_2^{(i)}, X_2^{(i)})}\right].$$

The expectation term on the right hand side is also finite by our assumptions.

Inductive Step: Now for the inductive step we will show that $\frac{\partial \theta_t^*}{\partial \theta_s^*}$ exists for all $s < t$ whenever $\frac{\partial \theta_{t-1}^*}{\partial \theta_{s'}^*}$ exists for all $s' < t-1$.

Recall that θ_t^* is defined as follows:

$$0 = \mathbb{E}_{\pi_{2:t}^*}\left[\psi_t(\mathcal{H}_t^{(i)}; \theta_t)\right].$$

We use weights $W_s^{(i)}(\theta^*_{s-1}, \hat{\theta}^{(n)}_{s-1}) \triangleq \frac{\pi_s(A_s^{(i)}, X_s^{(i)}; \theta^*_{s-1})}{\pi_s(A_s^{(i)}, X_s^{(i)}; \hat{\theta}^{(n)}_{s-1})}$. Using a similar argument as we did for the base case, for any $s < t$,

$$0 = \frac{\partial}{\partial \theta^*_s} \mathbb{E}_{\pi^*_{2:t}} \left[\psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \right] = \frac{\partial}{\partial \theta^*_s} \mathbb{E} \left[\left(\prod_{t'=2}^{t} W_{t'}^{(i)}(\theta^*_{t'-1}, \hat{\theta}^{(n)}_{t'-1}) \right) \psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \right]$$

$$= \sum_{s'=1}^{t-1} \mathbb{E} \left[\psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \left(\prod_{t'=2, t' \neq s'}^{t} W_{t'}^{(i)}(\theta^*_{t'-1}, \hat{\theta}^{(n)}_{t'-1}) \right) \left(\frac{\partial}{\partial \theta^*_s} W_{s'}^{(i)}(\theta^*_{s'-1}, \hat{\theta}^{(n)}_{s'-1}) \right) \right]$$
$$+ \mathbb{E} \left[\left(\prod_{t'=2}^{t} W_{t'}^{(i)}(\theta^*_{t'-1}, \hat{\theta}^{(n)}_{t'-1}) \right) \frac{\partial}{\partial \theta^*_s} \psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \right]$$

Note that for $s' < s$ that $\frac{\partial}{\partial \theta^*_s} W_{s'}^{(i)}(\theta^*_{s'-1}, \hat{\theta}^{(n)}_{s'-1}) = 0$. This is because we consider $\theta^*_{s'-1}$ to be an implicit function of $\theta^*_1, \theta^*_2, \ldots, \theta^*_{s'-2}$.

$$= \sum_{s'=s}^{t-1} \mathbb{E} \left[\psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \left(\prod_{t'=2, t' \neq s'}^{t} W_{t'}^{(i)}(\theta^*_{t'-1}, \hat{\theta}^{(n)}_{t'-1}) \right) \frac{\dot{\pi}_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)}; \theta^*_{s'-1})^\top}{\pi_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)}; \hat{\theta}^{(n)}_{s'-1})} \right] \frac{\partial \theta^*_{s'-1}}{\partial \theta^*_s}$$
$$+ \underbrace{\mathbb{E}_{\pi^*_{2:t}} \left[\frac{\partial}{\partial \theta^*_t} \psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \right]}_{= \dot{\Psi}_t} \frac{\partial \theta^*_t}{\partial \theta^*_s}$$

$$= \sum_{s'=s}^{t-1} \mathbb{E}_{\pi^*_{2:t}} \left[\psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \frac{\dot{\pi}_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)}; \theta^*_{s'-1})^\top}{\pi^*_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)})} \right] \frac{\partial \theta^*_{s'-1}}{\partial \theta^*_s} + \dot{\Psi}_t \frac{\partial \theta^*_t}{\partial \theta^*_s}$$

By invertibility of $\dot{\Psi}_t$ by assumption,

$$\frac{\partial \theta^*_t}{\partial \theta^*_s} = -\dot{\Psi}_t^{-1} \sum_{s'=s}^{t-1} \mathbb{E}_{\pi^*_{2:t}} \left[\psi_t(\mathcal{H}_t^{(i)}; \theta^*_t) \frac{\dot{\pi}_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)}; \theta^*_{s'-1})^\top}{\pi^*_{s'}(A_{s'}^{(i)}, X_{s'}^{(i)})} \right] \frac{\partial \theta^*_{s'-1}}{\partial \theta^*_s}$$

Note that all expectation terms in the statement above are finite by our assumptions and all the derivative terms $\frac{\partial \theta^*_{s'-1}}{\partial \theta^*_s}$ exist by the induction assumption and previous steps in the induction argument. ■

A.4. General Conditions that Can Replace Lipschitz Estimating Function Condition 4

Recall that $f_T(\,\cdot\,; \beta_{1:T-1}, \theta) \triangleq \big(\prod_{t=2}^{T} \pi_t(\,\cdot\,; \beta_{t-1})\big) \psi\big(\,\cdot\,; \theta\big)$ and the functions $f_t(\,\cdot\,; \beta_{1:t}) \triangleq \big(\prod_{s=2}^{t-1} \pi_s(\,\cdot\,; \beta_{s-1})\big) c_t^\top \phi_t\big(\,\cdot\,; \beta_t\big)$. Also recall we defined $\mathcal{F}_{T,c_T} \triangleq \big\{ c_T^\top f_T(\,\cdot\,; \beta_{1:T-1}, \theta) : \beta_{1:T-1} \in B_{T-1}, \theta \in \Theta \big\}$ for any fixed $c_T \in \mathbb{R}^{d_T}$. Similarly, we define $\mathcal{F}_{t,c_t} \triangleq \big\{ c_t^\top f_t(\,\cdot\,; \beta_{1:t-1}, \theta) : \beta_{1:t-1} \in B_{t-1} \big\}$ for any fixed $c_t \in \mathbb{R}^{d_t}$.

Condition 9 (Finite Bracketing Integral) *For each $t \in [1\colon T]$, we assume that for any finite, fixed $c_t \in \mathbb{R}^{d_t}$,*

$$\int_0^1 \sqrt{\log N_{[\,]}\left(\epsilon, \mathcal{F}_{t,c_t}, L_2(\mathcal{P}_{\pi^*_{2:t}})\right)} d\epsilon < \infty,$$

where $\mathcal{P}_{\pi^*_{2:t}}$ is the distribution of potential outcomes $\mathcal{P}$ where actions are selected with policies $\pi^*_{2:t}$. We also assume $\mathcal{F}_{t,c_t}$ has a measurable envelope function F_{t,c_t} with $\mathbb{E}_{\pi^*_{2:t}}\left[F_{t,c_t}(\mathcal{H}_t^{(i)})^2\right] < \infty$.

Recall that an envelope function F_{t,c_t} for the class of functions $\mathcal{F}_{t,c_t}$ means that $\sup_{f \in \mathcal{F}_{t,c_t}} |f(\mathcal{H}_t^{(i)})| < F_{t,c_t}(\mathcal{H}_t^{(i)}) < \infty$ w.p. 1 (Van der Vaart, 2000, pg. 270).

Condition 10 below ensures that the function class $\mathcal{F}_{T,c_T}$ is locally continuous in their parameters at $[\beta^*_{1:t}, \theta^*]$. We use the semi-metric $\rho_t(f, f') \triangleq \mathbb{E}_{\pi^*_{2:t}}\left[\left\|f(\mathcal{H}_t^{(i)}) - f'(\mathcal{H}_t^{(i)})\right\|^2\right]$ for any $f, f' \in \mathcal{F}_{t,c_t}$.

Condition 10 (Locally Continuous Function Classes) *For any $\epsilon > 0$, there exists a $\delta_{T,\epsilon} > 0$ such that for $\beta_{1:T-1} \in B_{1:T-1}$ and $\theta \in \Theta$ with $\|[\beta_{1:T-1}, \theta] - [\beta^*_{1:T-1}, \theta^*]\| < \delta_{T,\epsilon}$,*

$$\rho_T\left(f_T(\,\cdot\,; \beta_{1:T-1}, \theta), f_T(\,\cdot\,; \beta^*_{1:T-1}, \theta^*)\right) < \epsilon.$$

Similarly, for all $t \in [1\colon T-1]$, for any $\epsilon > 0$, there exists a $\delta_{t,\epsilon} > 0$ such that for $\beta_{1:t} \in \mathbb{R}^{\sum_{s=1}^t d_s}$ with $\|\beta_{1:t} - \beta^*_{1:t}\| < \delta_{t,\epsilon}$, then $\rho_t\left(f_t(\,\cdot\,; \beta_{1:t}), f_t(\,\cdot\,; \beta^*_{1:t})\right) < \epsilon$.

Lemma 5 (Condition 4 Implies Conditions 9 and 10 Hold) *Conditions 4 and 3 imply that Conditions 9 and 10 hold.*

Proof of Lemma 5:

Showing Condition 4 Holds: By Example 19.7 of Van der Vaart (2000), Condition 4 implies that $\int_0^1 \sqrt{\log N_{[\,]}\left(\epsilon, \mathcal{F}_{t,c_t}, L_2(\mathcal{P}_{\pi^*_{2:t}})\right)} d\epsilon < \infty$, where $\mathcal{P}_{\pi^*_{2:t}}$ is the distribution of potential outcomes $\mathcal{P}$ under policies $\pi^*_{2:t}$ for all $t \in [1\colon t]$.

We now show that there exists an envelope function F_{t,c_t} with $\mathbb{E}_{\pi^*_{2:t}}\left[F_{t,c_t}(\mathcal{H}_t^{(i)})^2\right] < \infty$ for all $t \in [1\colon T]$. Note that by Condition 4 for a non-negative function g_t with $\mathbb{E}_{\pi^*_{2:T}}\left[g_t\left(\mathcal{H}_t^{(i)}\right)^2\right] < \infty$,

$$\left\|f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right) - f_t\left(\mathcal{H}_t^{(i)}; \beta_{1:t}\right)\right\| \leq g_t\left(\mathcal{H}_t^{(i)}\right) \left\|\beta^*_{1:t} - \beta_{1:t}\right\|.$$

Thus we have that

$$\sup_{\beta_{1:t} \in B_{1:t}} \left\|c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right) - c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta_{1:t}\right)\right\| \leq \|c_t\| g_t\left(\mathcal{H}_t^{(i)}\right) \sup_{\beta_{1:t} \in B_{1:t}} \|\beta^*_{1:t} - \beta_{1:t}\|$$

Thus, for $F_{t,c_t}\left(\mathcal{H}_t^{(i)}\right) \triangleq \|c_t\| \left\|f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right)\right\| + \|c_t\| g_t\left(\mathcal{H}_t^{(i)}\right) \sup_{\beta_{1:t} \in B_{1:t}} \|\beta^*_{1:t} - \beta_{1:t}\|$, we have that for any $f_t(\mathcal{H}_t^{(i)}; \beta_{1:t})$ with $\beta_{1:t} \in B_{1:t}$,

$$\begin{aligned}
c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta_{1:t}\right) &\leq \left\|c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta_{1:t}\right)\right\| \\
&\leq \left\|c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right)\right\| + \left\|c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right) - c_t^\top f_t\left(\mathcal{H}_t^{(i)}; \beta_{1:t}\right)\right\| \\
&\leq \|c_t\| \left\|f_t\left(\mathcal{H}_t^{(i)}; \beta^*_{1:t}\right)\right\| + \|c_t\| g_t\left(\mathcal{H}_t^{(i)}\right) \left\|\beta^*_{1:t} - \beta_{1:t}\right\| \leq F_{t,c_t}\left(\mathcal{H}_t^{(i)}\right).
\end{aligned}$$

We now show that the second moment of $F_{t,c_t}\big(\mathcal{H}_t^{(i)}\big)$ is bounded:

$$\mathbb{E}_{\pi^*_{2:t}}\left[F_{t,c_t}\big(\mathcal{H}_t^{(i)}\big)^2\right] = \mathbb{E}_{\pi^*_{2:t}}\left[\left\{\|c_t\|\left\|f_t\big(\mathcal{H}_t^{(i)};\beta^*_{1:t}\big)\right\| + \|c_t\|g_t\big(\mathcal{H}_t^{(i)}\big)\sup_{\beta_{1:t}\in B_{1:t}}\|\beta^*_{1:t}-\beta_{1:t}\|\right\}^2\right]$$

$$\leq 3\|c_t\|^2\mathbb{E}_{\pi^*_{2:t}}\left[\left\|f_t\big(\mathcal{H}_t^{(i)};\beta^*_{1:t}\big)\right\|^2\right] + 3\|c_t\|^2\mathbb{E}_{\pi^*_{2:t}}\left[g_t\big(\mathcal{H}_t^{(i)}\big)^2\right]\left\{\sup_{\beta_{1:t}\in B_{1:t}}\|\beta^*_{1:t}-\beta_{1:t}\|\right\}^2.$$

Above, $\mathbb{E}_{\pi^*_{2:t}}\left[\left\|f_t\big(\mathcal{H}_t^{(i)};\beta^*_{1:t}\big)\right\|^2\right] < \infty$ by Condition 3, $\mathbb{E}_{\pi^*_{2:t}}\left[g_t\big(\mathcal{H}_t^{(i)}\big)^2\right] < \infty$ by Condition 4, and $\sup_{\beta_{1:t}\in B_{1:t}}\|\beta^*_{1:t}-\beta_{1:t}\|$ is bounded by our assumption that B_t are bounded.

Showing Condition 10 Holds: Let $t \in [1\colon T]$ and $\epsilon > 0$.

$$\rho_t\Big(f_t(\,\cdot\,;\beta_{1:t}), f_t(\,\cdot\,;\beta^*_{1:t})\Big) = \mathbb{E}_{\pi^*_{2:t}}\left[\left\|f_t(\mathcal{H}_t^{(i)};\beta^*_{1:t}) - f_t(\mathcal{H}_t^{(i)};\beta_{1:t})\right\|^2\right]$$

By Condition 4, $\left\|f_t(\mathcal{H}_t^{(i)};\beta^*_{1:t}) - f_t(\mathcal{H}_t^{(i)};\beta_{1:t})\right\| \leq g_t(\mathcal{H}_t^{(i)})\,\|\beta^*_{1:t}-\beta_{1:t}\|$, for a non-negative function g_t with $\mathbb{E}_{\pi^*_{2:T}}\big[g_t\big(\mathcal{H}_t^{(i)}\big)^2\big] < \infty$ so

$$\leq \mathbb{E}_{\pi^*_{2:t}}\left[g_t(\mathcal{H}_t^{(i)})^2\right]\|\beta^*_{1:t}-\beta_{1:t}\|^2.$$

Thus, if $\|\beta^*_{1:t}-\beta_{1:t}\| < \mathbb{E}_{\pi^*_{2:t}}\left[g_t(\mathcal{H}_t^{(i)})^2\right]^{-1/2}\sqrt{\epsilon}$, then $\rho_t\big(f_t(\,\cdot\,;\beta_{1:t}), f_t(\,\cdot\,;\beta^*_{1:t})\big) < \epsilon$. ■

Appendix B. Consistency

Theorem 1 (Consistency) *We consider the setting of Section 1. Under Conditions 1, 3, 5, 6, 7, and 9, $\hat{\theta}^{(n)} \xrightarrow{P} \theta^*$ and $\hat{\beta}_t^{(n)} \xrightarrow{P} \beta_t^*$ for all $t \in [1\colon T-1]$. Moreover, $\|\hat{\theta}^{(n)}(\,\cdot\,) - \theta^*(\,\cdot\,)\|_{B_{1:T-1}} \xrightarrow{P} 0$ and $\|\hat{\beta}_t^{(n)}(\,\cdot\,) - \beta_t^*(\,\cdot\,)\|_{B_{1:-1}} \xrightarrow{P} 0$ for all $t \in [1\colon T-1]$.*

Note above our Theorem statement differs from that in the main text because we do not use Condition 4 and instead use Condition 9; see Lemma 5 for a proof that Condition 4 implies that Condition 9 holds.

Condition 1 (Minimum Exploration) *For some $\pi_{\min} > 0$, for all $t \in [1\colon T]$,*

$$\min_{a \in \mathcal{A}} \hat{\pi}_t\big(a, X_t^{(i)}\big) \geq \pi_{\min} \text{ w.p. } 1 \quad \text{and} \quad \inf_{\beta_{t-1} \in B_{t-1}} \min_{a \in \mathcal{A}} \pi_t\big(a, X_t^{(i)}; \beta_{t-1}\big) \geq \pi_{\min} \text{ w.p. } 1.$$

Condition 3 (Finite Moments) *For some $\alpha > 0$, for all $t \in [1\colon T]$,*

$$\mathbb{E}_{\pi^*_{2:t}}\left[\left\|\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\right\|_1^{4+\alpha}\right] < \infty \quad \text{and} \quad \mathbb{E}_{\pi^*_{2:T}}\left[\left\|\psi_T\big(\mathcal{H}_T^{(i)}; \theta^*\big)\right\|_1^{4+\alpha}\right] < \infty.$$

Condition 4 (Lipschitz Estimating Functions) *There exists a function $g_T\big(\mathcal{H}_T^{(i)}\big)$ such that (i) $\mathbb{E}_{\pi^*_{2:T}}\big[g_T\big(\mathcal{H}_T^{(i)}\big)^2\big] < \infty$ and (ii) for all $\beta_{1:T-1}, \beta'_{1:T-1} \in B_{1:T-1}$, $\theta, \theta' \in \Theta$,*

$$\big\|f_T(\mathcal{H}_T^{(i)}; \beta_{1:T-1}, \theta) - f_T(\mathcal{H}_T^{(i)}; \beta'_{1:T-1}, \theta')\big\| \leq g_T\big(\mathcal{H}_T^{(i)}\big)\big\|[\beta_{1:T-1}, \theta] - [\beta'_{1:T-1}, \theta']\big\|.$$

Also for all $t \in [2\colon T]$, there exists a function $g_t\big(\mathcal{H}_t^{(i)}\big)$ such that (i) $\mathbb{E}_{\pi^*_{2:t}}\big[g_t\big(\mathcal{H}_t^{(i)}\big)^2\big] < \infty$ and (ii) for all $\beta_{1:t}, \beta'_{1:t} \in B_{1:t}$, $\|f_t(\mathcal{H}_t^{(i)}; \beta_{1:t}, \theta) - f_t(\mathcal{H}_t^{(i)}; \beta_{1:t})\| \leq g_t(\mathcal{H}_t^{(i)})\|\beta_{1:t} - \beta'_{1:t}\|$.

Condition 5 (Fréchet Differentiability) *$\theta^{*,[s-1]}(\,\cdot\,)$ is Fréchet differentiable with respect to $\beta_s^*(\,\cdot\,)$ and continuous in $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ for all $s < T$, and for all $t \in [1\colon T-1]$, $\beta_t^{*,[s-1]}(\,\cdot\,)$ is Fréchet differentiable with respect to $\beta_s^*(\,\cdot\,)$ and continuous in $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ for all $s < t$. Also, the derivative functions $\frac{\partial \theta^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ and $\frac{\partial \beta_t^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ are continuous in their arguments $\beta_{1:s-1} \in B_{1:s-1}$. $\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ is Fréchet differentiable with respect to $\theta^*(\,\cdot\,)$, and $\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ is Fréchet differentiable with respect to $\beta_t^*(\,\cdot\,)$ for all $t \in [1\colon T-1]$. Also, derivative functions $\frac{\partial}{\partial \theta^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ and $\frac{\partial}{\partial \beta_t^*(\,\cdot\,)}\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ are continuous in their arguments, $\beta_{1:T-1} \in B_{1:T-1}$ and $\beta_{1:t-1} \in B_{1:t-1}$, respectively.*

Condition 6 (Well-Separated Solution) *For any $\epsilon > 0$, there exists some $\eta_\epsilon > 0$ such that for all $\beta_{1:T-1} \in B_{1:T-1}$,*

$$\inf_{\theta \in \mathbb{R}^{d_T} : \|\theta - \theta^*(\beta_{1:T-1})\| > \epsilon} \left\|\mathbb{E}_{\pi_2(\beta_1), \pi_3(\beta_2), \ldots, \pi_T(\beta_{T-1})}\big[\psi_T(\mathcal{H}_T^{(i)}; \theta)\big]\right\| > \eta_\epsilon.$$

Similarly, for $t \in [1\colon T-1]$, for any $\epsilon > 0$ there exists some $\eta_{t,\epsilon} > 0$ such that for all $\beta_{1:t-1} \in B_{1:t-1}$, $\inf_{\beta_t \in \mathbb{R}^{d_t} : \|\beta_t - \beta_t^*(\beta_{1:t-1})\| > \epsilon} \left\|\mathbb{E}_{\pi_2(\beta_1), \pi_3(\beta_2), \ldots, \pi_t(\beta_{t-1})}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t\big)\big]\right\| > \eta_{t,\epsilon}$.

Condition 7 (Compact Parameter Space In Probability)
$\mathbb{P}\big(\{\hat{\theta}^{(n)}(\beta_{1:T-1}) : \beta_{1:T-1} \in B_{1:T-1}\} \subset \Theta\big) \to 1$ *and*
$\mathbb{P}\big(\{\hat{\beta}_t^{(n)}(\beta_{1:t-1}) : \beta_{1:t-1} \in B_{1:t-1}\} \subset B_t\big) \to 1$ *for all* $t \in [1\colon T-1]$.

Recall that $f_T(\,\cdot\,;\beta_{1:T-1},\theta) \triangleq \big(\prod_{t=2}^{T}\pi_t(\,\cdot\,;\beta_{t-1})\big)\psi(\,\cdot\,;\theta)$ and the functions $f_t(\,\cdot\,;\beta_{1:t}) \triangleq \big(\prod_{s=2}^{t-1}\pi_s(\,\cdot\,;\beta_{s-1})\big)\phi_t(\,\cdot\,;\beta_t)$. Also recall we defined $\mathcal{F}_{T,c_T} \triangleq \{c_T^\top f_T(\,\cdot\,;\beta_{1:T-1},\theta) : \beta_{1:T-1} \in B_{T-1}, \theta \in \Theta\}$ for any fixed $c_T \in \mathbb{R}^{d_T}$. Similarly, we define $\mathcal{F}_{t,c_t} \triangleq \{c_t^\top f_t(\,\cdot\,;\beta_{1:t-1},\theta) : \beta_{1:t-1} \in B_{t-1}\}$ for any fixed $c_t \in \mathbb{R}^{d_t}$.

Condition 9 (Finite Bracketing Integral) *For each* $t \in [1\colon T]$*, we assume that for any finite, fixed* $c_t \in \mathbb{R}^{d_t}$*,*

$$\int_0^1 \sqrt{\log N_{[\,]}\left(\epsilon, \mathcal{F}_{t,c_t}, L_2\big(\mathcal{P}_{\pi^*_{2:t}}\big)\right)}d\epsilon < \infty,$$

where $\mathcal{P}_{\pi^*_{2:t}}$ *is the distribution of potential outcomes* $\mathcal{P}$ *where actions are selected with policies* $\pi^*_{2:t}$*. We also assume* $\mathcal{F}_{t,c_t}$ *has a measurable envelope function* F_{t,c_t} *with* $\mathbb{E}_{\pi^*_{2:t}}\big[F_{t,c_t}(\mathcal{H}_t^{(i)})^2\big] < \infty$.

Recall that an envelope function F_{t,c_t} for the class of functions $\mathcal{F}_{t,c_t}$ means that $\sup_{f\in\mathcal{F}_{t,c_t}}|f(\mathcal{H}_t^{(i)})| < F_{t,c_t}(\mathcal{H}_t^{(i)}) < \infty$ w.p. 1 (Van der Vaart, 2000, pg. 270).

Notation for the Remainder of Section (Appendix B) For notational convenience, we let $\theta_T \triangleq \theta$ and $\theta_t \triangleq \beta_t$ for all $t \in [1\colon T-1]$. This means that $\theta_T^* \triangleq \theta^*$, $\theta_t^* \triangleq \beta_t^*$, $\hat{\theta}_T^{(n)} \triangleq \hat{\theta}^{(n)}$, and $\hat{\theta}_t^{(n)} \triangleq \hat{\beta}_t^{(n)}$. Also we use $\Theta_T \triangleq \Theta$ and $\Theta_t \triangleq B_t$, where recall Θ is a bounded ball that contains $\theta^*(\beta_{1:T-1})$ for all $\beta_{1:T-1} \in B_{1:T-1}$ and B_t is a bounded ball that contains $\beta_t^*(\beta_{1:t-1})$ for all $\beta_{1:t-1} \in B_{1:t-1}$. Additionally, we let $\psi_T \triangleq \psi$ and $\psi_t \triangleq \phi_t$ for all $t \in [1\colon T-1]$, so $\psi_T(\mathcal{H}_T^{(i)};\theta_T) = \psi(\mathcal{H}_T^{(i)};\theta)$ and $\psi_t(\mathcal{H}_t^{(i)};\theta_t) = \phi_t(\mathcal{H}_t^{(i)};\beta_t)$ for $t < T$. We also define the following notation:

$$W_{2:t}^{(i)}(\theta_{1:t-1},\hat{\theta}_{1:t-1}^{(n)}) \triangleq \prod_{s=2}^{t} W_s^{(i)}(\theta_{s-1},\hat{\theta}_{s-1}^{(n)})$$

$$\Psi_t(\theta_{1:t}) \triangleq \mathbb{E}_{\pi_2(\theta_1),\pi_3(\theta_2),\ldots,\pi_t(\theta_{t-1})}\big[\psi_t(\mathcal{H}_t^{(i)};\theta_t)\big] = \mathbb{E}\left[W_{2:t}^{(i)}(\theta_{1:t-1},\hat{\theta}_{1:t-1}^{(n)})\psi_t(\mathcal{H}_t^{(i)};\theta_t)\right]$$

$$\hat{\Psi}_t^{(n)}(\theta_{1:t}) \triangleq \frac{1}{n}\sum_{i=1}^{n} W_{2:t}^{(i)}(\theta_{1:t-1},\hat{\theta}_{1:t-1}^{(n)})\psi_t(\mathcal{H}_t^{(i)};\theta_t)$$

B.1. Proof of Theorem 1

$t = 1$ **Case (Base Case):** We first show that $\hat{\theta}_1^{(n)} \xrightarrow{P} \theta_1^*$. We follow an argument similar to that of Van der Vaart (2000, Theorem 5.7).

Let $\epsilon > 0$. By the well-separated solution Condition 6, for some $\eta_\epsilon > 0$,

$$\mathbb{P}\left(\|\hat{\theta}_1^{(n)} - \theta_1^*\| > \epsilon\right) \le \mathbb{P}\left(\left\|\mathbb{E}\left[\psi_1(\mathcal{H}_1^{(i)};\hat{\theta}_1^{(n)})\right]\right\| > \eta_\epsilon\right) = \mathbb{P}\left(\left\|\Psi_1(\hat{\theta}_1^{(n)})\right\| > \eta_\epsilon\right).$$

Thus it is sufficient to show that $\left\|\Psi_1(\hat{\theta}_1^{(n)})\right\| = o_P(1)$. By triangle inequality

$$\left\|\Psi_1(\hat{\theta}_1^{(n)})\right\| \le \left\|\Psi_1(\hat{\theta}_1^{(n)}) - \hat{\Psi}_1^{(n)}(\hat{\theta}_1^{(n)})\right\| + \left\|\hat{\Psi}_1^{(n)}(\hat{\theta}_1^{(n)})\right\|$$

Note that $\hat{\Psi}_1^{(n)}\big(\hat{\theta}_1^{(n)}\big) = 0$ by definition of $\hat{\theta}_1^{(n)}$. By Condition 7, $\mathbb{P}\big(\hat{\theta}_1^{(n)} \in \Theta_1\big) \to 1$, so

$$= o_P(1) + \mathbb{I}_{\hat{\theta}_1^{(n)} \in \Theta_1} \left\| \Psi_1\big(\hat{\theta}_1^{(n)}\big) - \hat{\Psi}_1^{(n)}\big(\hat{\theta}_1^{(n)}\big) \right\|$$

$$\leq o_P(1) + \sup_{\theta_1 \in \Theta_1} \left\| \Psi_1(\theta_1) - \hat{\Psi}_1^{(n)}(\theta_1) \right\| \xrightarrow{P} 0.$$

By our bracketing complexity assumption (Condition 9), the above limit holds by the uniform law of large numbers for i.i.d. random variables (Van der Vaart, 2000, Theorem 19.4).

General t case (Induction Step): For our induction assumption, we assume that $\hat{\theta}_s^{(n)} \xrightarrow{P} \theta_s^*$ and $\sup_{\theta_{1:s-1} \in \Theta_{1:s-1}} \left\| \hat{\theta}_s^{(n)}(\theta_{1:s-1}) - \theta_s^*(\theta_{1:s-1}) \right\| \xrightarrow{P} 0$ for all $s \in [1\colon t-1]$. We will show that $\hat{\theta}_t^{(n)} \xrightarrow{P} \theta_t^*$ and $\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \xrightarrow{P} 0$.

Let $\epsilon > 0$.

$$\mathbb{P}\left(\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| > \epsilon \right)$$

We use $\theta_{1:t-1} \in \Theta_{1:t-1}$ to denote $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$, $\ldots$, $\theta_{t-1} \in \Theta_{1:t-1}$. By the well-separated solution Condition 6, for some $\eta_\epsilon > 0$,

$$\leq \mathbb{P}\left(\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \mathbb{E}_{\pi_2(\theta_1), \pi_3(\theta_2), \ldots, \pi_t(\theta_{t-1})} \left[\psi_t\big(\mathcal{H}_t^{(i)}; \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) \right] \right\| > \eta_\epsilon \right)$$

$$= \mathbb{P}\left(\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\| > \eta_\epsilon \right)$$

Thus, it is sufficient to show that $\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\| = o_P(1)$. By triangle inequality,

$$\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\|$$

$$\leq \sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) - \hat{\Psi}_t^{(n)}\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\|$$

$$+ \underbrace{\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \hat{\Psi}_t^{(n)}\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\|}_{=0}$$

The term on the second line above equals zero by the definition of $\hat{\theta}_t^{(n)}(\,\cdot\,)$.

By Condition 7, $\mathbb{P}\left(\left\{ \hat{\theta}_t^{(n)}(\theta_{1:t-1}) : \theta_{1:t-1} \in \Theta_{1:t-1} \right\} \subset \Theta_t \right) \to 1$, thus,
$\mathbb{I}_{\hat{\theta}_t^{(n)}(\,\cdot\,) \in \Theta_t} \triangleq \mathbb{I}_{\left\{ \hat{\theta}_t^{(n)}(\theta_{1:t-1}) : \theta_{1:t-1} \in \Theta_{1:t-1} \right\}} \xrightarrow{P} 1$,

$$= o_P(1) + \mathbb{I}_{\hat{\theta}_t^{(n)}(\,\cdot\,) \in \Theta_t} \sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) - \hat{\Psi}_t^{(n)}\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) \right\|$$

$$\leq o_P(1) + \sup_{\theta_{1:t} \in \Theta_{1:t}} \left\| \hat{\Psi}_t^{(n)}(\theta_{1:t}) - \Psi_t(\theta_{1:t}) \right\| \xrightarrow{P} 0.$$

The above limit holds because by Conditions 1, 3, and 9, we can apply Lemma 6 to show that the above converges in probability to 0 (see Appendix B.2). Thus, we have that

$$\sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \xrightarrow{P} 0. \tag{19}$$

We now show that $\hat{\theta}_t^{(n)} \xrightarrow{P} \theta_t^*$, i.e., $\hat{\theta}_t^{(n)}\left(\hat{\theta}_{1:t-1}^{(n)}\right) \xrightarrow{P} \theta_t^*\left(\theta_{1:t-1}^*\right)$. By triangle inequality,

$$\left\| \hat{\theta}_t^{(n)}\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^*\left(\theta_{1:t-1}^*\right) \right\| \leq \left\| \hat{\theta}_t^{(n)}\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^*\left(\hat{\theta}_{1:t-1}^{(n)}\right) \right\| + \left\| \theta_t^*\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^*\left(\theta_{1:t-1}^*\right) \right\|$$

We show that both of the terms on the right hand side above are $o_P(1)$.

First Term By our induction assumption $\hat{\theta}_{1:t-1}^{(n)} \xrightarrow{P} \theta_{1:t-1}^*$, so $\mathbb{I}_{\hat{\theta}_{1:t-1}^{(n)} \in \Theta_{1:t-1}} \xrightarrow{P} 1$.

$$\left\| \hat{\theta}_t^{(n)}\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^*\left(\hat{\theta}_{1:t-1}^{(n)}\right) \right\| = o_P(1) + \mathbb{I}_{\hat{\theta}_{1:t-1}^{(n)} \in \Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^*\left(\hat{\theta}_{1:t-1}^{(n)}\right) \right\|$$

$$\leq o_P(1) + \sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \xrightarrow{P} 0.$$

The final limit holds by Equation (19).

Second Term By Condition 5, $\theta_t^{*,[t-1]}(\,\cdot\,)$ is continuous in $\theta_{t-1}(\,\cdot\,) : \Theta_{1:t-2} \mapsto \Theta_{t-1}$. By our induction assumption, $\sup_{\theta_{1:t-2} \in \Theta_{1:t-2}} \left\| \hat{\theta}_{t-1}^{(n)}(\theta_{1:t-2}) - \theta_{t-1}^*(\theta_{1:t-2}) \right\| \xrightarrow{P} 0$, so by continuous mapping theorem,

$$\sup_{\theta_{1:t-2} \in \Theta_{1:t-2}} \left\| \theta_t^{*,[t-1]}\left(\theta_{1:t-2}, \hat{\theta}_{t-1}^{(n)}(\theta_{1:t-2})\right) - \underbrace{\theta_t^{*,[t-1]}\left(\theta_{1:t-2}, \theta_{t-1}^*(\theta_{1:t-2})\right)}_{=\theta_t^{*,[t-2]}(\theta_{1:t-2})} \right\| \xrightarrow{P} 0.$$

By our induction assumption $\hat{\theta}_{1:t-2}^{(n)} \xrightarrow{P} \theta_{1:t-2}^*$, so $\mathbb{P}\left(\hat{\theta}_{1:t-2}^{(n)} \in \Theta_{1:t-2}\right) \to 1$. Thus, the above result implies that $\left\| \theta_t^*\left(\hat{\theta}_{1:t-1}^{(n)}\right) - \theta_t^{*,[t-2]}\left(\hat{\theta}_{1:t-2}^{(n)}\right) \right\| \xrightarrow{P} 0$.

Applying a similar argument again, by Condition 5, $\theta_t^{*,[t-2]}(\,\cdot\,)$ is continuous in $\theta_{t-2}(\,\cdot\,) : \Theta_{1:t-3} \mapsto \Theta_{t-2}$.By our induction assumption, $\sup_{\theta_{1:t-3} \in \Theta_{1:t-3}} \left\| \hat{\theta}_{t-2}^{(n)}(\theta_{1:t-3}) - \theta_{t-2}^*(\theta_{1:t-3}) \right\| \xrightarrow{P} 0$, so by continuous mapping theorem,

$$\sup_{\theta_{1:t-3} \in \Theta_{1:t-3}} \left\| \theta_t^{*,[t-2]}\left(\theta_{1:t-3}, \hat{\theta}_{t-2}^{(n)}(\theta_{1:t-3})\right) - \underbrace{\theta_t^{*,[t-2]}\left(\theta_{1:t-3}, \theta_{t-2}^*(\theta_{1:t-3})\right)}_{=\theta_t^{*,[t-3]}(\theta_{1:t-3})} \right\| \xrightarrow{P} 0.$$

By our induction assumption $\hat{\theta}_{1:t-3}^{(n)} \xrightarrow{P} \theta_{1:t-3}^*$, so $\mathbb{P}\left(\hat{\theta}_{1:t-3}^{(n)} \in \Theta_{1:t-3}\right) \to 1$. Thus, the above result implies that $\left\| \theta_t^{*,[t-2]}\left(\hat{\theta}_{1:t-2}^{(n)}\right) - \theta_t^{*,[t-3]}\left(\hat{\theta}_{1:t-3}^{(n)}\right) \right\| \xrightarrow{P} 0$.

By repeatedly applying the above argument we eventually get that for all $s \in [1\colon t-1]$,

$$\sup_{\theta_{1:s-1}\in\Theta_{1:s-1}} \left\| \theta_t^{*,[s]}\big(\theta_{1:s-1}, \hat{\theta}_s^{(n)}(\theta_{1:s-1})\big) - \underbrace{\theta_t^{*,[s]}\big(\theta_{1:s-1}, \theta_s^*(\theta_{1:s-1})\big)}_{=\theta_t^{*,[s-1]}(\theta_{1:s-1})} \right\| \xrightarrow{P} 0.$$

and that

$$\left\| \theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big) - \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big) \right\| \xrightarrow{P} 0.$$

Thus, the desired result holds because by telescoping series and triangle inequality,

$$\left\| \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\left(\theta_{1:t-1}^*\right) \right\| \leq \sum_{s=1}^{t-1} \left\| \theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big) - \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big) \right\| \xrightarrow{P} 0.$$

Note that for $s = 1$, $\theta_t^*\left(\theta_{1:t-1}^*\right) = \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big)$.
Also for $s = t-1$, $\theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big) = \theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big)$. ■

B.2. Lemma 6: Importance-Weighted Uniform Weak Law of Large Numbers

Lemma 6 (Weighted Martingale Weak Uniform Law of Large Numbers) *We consider the problem setting as described in Section 1. Under Conditions 1, 3, and 9,*

$$\sup_{\theta_{1:t}\in\Theta_{1:t}} \left\| \hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big) - \Psi_t\big(\theta_{1:t}\big) \right\| \xrightarrow{P} 0.$$

Proof for Theorem 6 It is sufficient to show that for any $c_t \in \mathbb{R}^{d_t}$ that the following converges in probability to 0:

$$\sup_{\theta_{1:t}\in\Theta_{1:t}} \left| c_t^\top \hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big) - c_t^\top \Psi_t\big(\theta_{1:t}\big) \right|$$

$$= \sup_{\theta_{1:t}\in\Theta_{1:t}} \left| \frac{1}{n}\sum_{i=1}^{n} \left\{ W_{2:t}^{(i)}(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}) c_t^\top \psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big) \right.\right.$$
$$\left.\left. - \mathbb{E}\left[W_{2:t}^{(i)}(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}) c_t^\top \psi_t(\mathcal{H}_t^{(i)}; \theta_t) \right] \right\} \right|.$$

Note that for the class of functions $\mathcal{F}_{t,c_t}$ defined above Condition 9,

$$\left\{ W_t^{(i)}\big(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}\big) c_t^\top \psi_t(\mathcal{H}_t^{(i)}; \theta_t) : \theta_{1:t} \in \Theta_{1:t} \right\} = \left\{ \big(\hat{\pi}_{2:t}^{(i)}\big)^{-1} f\big(\mathcal{H}_t^{(i)}\big) : f \in \mathcal{F}_{t,c_t} \right\}, \quad (20)$$

where $\hat{\pi}_s^{(i)} \triangleq \hat{\pi}_s\big(A_s^{(i)}, X_s^{(i)}\big)$ and $\hat{\pi}_{2:t}^{(i)} = \prod_{s=2}^{t} \hat{\pi}_s^{(i)}$.

We now follow an argument similar to that for (Van Der Vaart and Wellner, 1996, Theorem 2.4.1). By Condition 9, we have that for any $\epsilon > 0$, $N_{[\,]}\left(\epsilon, \mathcal{F}_{t,c_t}, L_2\big(\mathcal{P}_{\pi_{2:t}^*}\big)\right) < \infty$. This means that we can find finitely many brackets $[l_k, u_k]$ that cover $\mathcal{F}_{t,c_t}$ with $\mathbb{E}_{\pi_{2:t}^*}\left[\big(u_k(\mathcal{H}_t^{(i)}) - l_k(\mathcal{H}_t^{(i)})\big)^2\right]^{1/2} < \epsilon$. So for

any $f \in \mathcal{F}_{t,c_t}$, we can find a bracket $[l_k, u_k]$ that contains f, i.e., $l_k(\mathcal{H}_t^{(i)}) \leq f(\mathcal{H}_t^{(i)}) \leq u_k(\mathcal{H}_t^{(i)})$. So,

$$\frac{1}{n}\sum_{i=1}^{n}\left\{(\hat{\pi}_{2:t}^{(i)})^{-1} f(\mathcal{H}_t^{(i)}) - \mathbb{E}\left[(\hat{\pi}_{2:t}^{(i)})^{-1} f(\mathcal{H}_t^{(i)})\right]\right\} \tag{21}$$

$$\leq \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left\{(\hat{\pi}_{2:t}^{(i)})^{-1} u_k(\mathcal{H}_t^{(i)}) - \mathbb{E}\left[(\hat{\pi}_{2:t}^{(i)})^{-1} u_k(\mathcal{H}_t^{(i)})\right]\right\}}_{=(a)}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(\hat{\pi}_{2:t}^{(i)})^{-1}\left(u_k(\mathcal{H}_t^{(i)}) - f(\mathcal{H}_t^{(i)})\right)\right]}_{=(b)}$$

Note we can upper bound term (a) by the worst case upper bracket out of the $N_{[\,]}\left(\epsilon, \mathcal{F}_{t,c_t}, L_2(\mathcal{P}_{\pi_{2:t}^*})\right) < \infty$ brackets:

$$\max_k \frac{1}{n}\sum_{i=1}^{n}\left\{(\hat{\pi}_{2:t}^{(i)})^{-1} u_k(\mathcal{H}_t^{(i)}) - \mathbb{E}\left[(\hat{\pi}_{2:t}^{(i)})^{-1} u_k(\mathcal{H}_t^{(i)})\right]\right\} = o_P(1).$$

The limit above holds by Lemma 7 (Importance-Weighted Weak Law of Large Numbers result) because there are finitely many brackets; note that when applying Lemma 7 we set $f_t(\mathcal{H}_t^{(i)}) = \left(\prod_{s=2}^{t}\pi_s^*(A_s^{(i)}, X_s^{(i)})\right)^{-1} u_k(\mathcal{H}_t^{(i)})$ and $f_s(\mathcal{H}_s^{(i)}) = 0$ for all $s \neq t$ and
$\mathbb{E}_{\pi_{2:t}^*}\left[\left(\prod_{s=2}^{t}\pi_s^*(A_s^{(i)}, X_s^{(i)})\right)^{-2} u_k(\mathcal{H}_t^{(i)})^2\right] \leq \pi_{\min}^{-2(t-1)}\mathbb{E}_{\pi_{2:t}^*}\left[u_k(\mathcal{H}_t^{(i)})^2\right]$
$\leq \pi_{\min}^{-2(t-1)}\mathbb{E}_{\pi_{2:t}^*}\left[F_{t,c_t}(\mathcal{H}_t^{(i)})^2\right] < \infty$ by Conditions 1 and 9.

Note, since $(\hat{\pi}_{2:t}^{(i)})^{-1} \leq \pi_{\min}^{-(t-1)}$ and $\pi_{\min}^{t-1} \leq \left(\frac{\pi_{\min}}{1-\pi_{\min}}\right)^{t-1} \leq W_{2:t}^{(i)}(\theta_{1:t-1}^*, \hat{\theta}_{1:t-1})$ w.p. 1, so $(\hat{\pi}_{2:t}^{(i)})^{-1} \leq \pi_{\min}^{-(t-1)} \leq \pi_{\min}^{-2(t-1)} W_{2:t}^{(i)}(\theta_{1:t-1}^*, \hat{\theta}_{1:t-1})$. Thus, we can upper bound term (b):

$$\pi_{\min}^{-2(t-1)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t}^{(i)}(\theta_{1:t-1}^*, \hat{\theta}_{1:t-1})\left(u_k(\mathcal{H}_t^{(i)}) - f(\mathcal{H}_t^{(i)})\right)\right]$$

$$= \pi_{\min}^{-2(t-1)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\pi_{2:t}^*}\left[u_k(\mathcal{H}_t^{(i)}) - f(\mathcal{H}_t^{(i)})\right] \leq \pi_{\min}^{-2(t-1)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\pi_{2:t}^*}\left[u_k(\mathcal{H}_t^{(i)}) - l_k(\mathcal{H}_t^{(i)})\right]$$

$$\leq \pi_{\min}^{-2(t-1)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\pi_{2:t}^*}\left[\left\{u_k(\mathcal{H}_t^{(i)}) - l_k(\mathcal{H}_t^{(i)})\right\}^2\right]^{1/2} < \pi_{\min}^{-2(t-1)}\epsilon.$$

Recall, that since $\pi_{\min}^{-2(t-1)}$ is a constant, the above implies that we can make term (b) arbitrarily small.

We can make a similar argument for a lower bound for Equation (21). ■

B.3. Lemma 7: Importance-Weighted Weak Law of Large Numbers

Lemma 7 (Importance-Weighted Weak Law of Large Numbers) *We consider the problem setting as described in Section 1. Let $f_1, f_2, f_3, ..., f_T$ be real-valued functions of $\mathcal{H}_1^{(i)}, \mathcal{H}_2^{(i)}, \mathcal{H}_3^{(i)}, ..., \mathcal{H}_T^{(i)}$ respectively such that $\mathbb{E}_{\pi^*_{2:t}}\left[f_t\left(\mathcal{H}_t^{(i)}\right)^2\right] < \infty$ for all $t \in [1\colon T]$. We show that under Condition 1, for any $\{\theta_t\}_{t=1}^T$ such that $\theta_t \in \Theta_t$,*

$$\frac{1}{n}\sum_{i=1}^{n}\left\{f_1(\mathcal{H}_1^{(i)}) + \sum_{t=2}^{T} W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)})\right\} \xrightarrow{P} \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right].$$

Proof of Lemma 7 Note that by the weak law of large numbers for independent random variables, $\frac{1}{n}\sum_{i=1}^{n} f_1(\mathcal{H}_1^{(i)}) \xrightarrow{P} \mathbb{E}\left[f_1(\mathcal{H}_1^{(i)})\right]$. By Slutsky's Theorem, it is sufficient to show that for all $t \in [2\colon T]$,

$$\frac{1}{n}\sum_{i=1}^{n} W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)}) \xrightarrow{P} \mathbb{E}_{\pi^*_{2:T}}\left[f_t(\mathcal{H}_t^{(i)})\right].$$

Let $\epsilon > 0$. It is sufficient to show that the following converges to zero:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left\{W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)}) - \mathbb{E}_{\pi^*_{2:T}}\left[f_t(\mathcal{H}_t^{(i)})\right]\right\}\right| > \epsilon\right)$$

For convenience, let $\tilde{f}_t(\mathcal{H}_t^{(i)}) \triangleq W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)}) - \mathbb{E}_{\pi^*_{2:T}}\left[f_t(\mathcal{H}_t^{(i)})\right]$. It is sufficient to show that By Markov inequality,

$$= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{f}_t(\mathcal{H}_t^{(i)})\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\mathbb{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{f}_t(\mathcal{H}_t^{(i)})\right\}^2\right]$$

$$= \underbrace{\frac{1}{\epsilon^2 n^2}\sum_{i=1}^{n}\mathbb{E}\left[\tilde{f}_t(\mathcal{H}_t^{(i)})^2\right]}_{=(a)} + \underbrace{\frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\left[\tilde{f}_t(\mathcal{H}_t^{(i)})\tilde{f}_t(\mathcal{H}_t^{(j)})\right]}_{=(b)}$$

We now show that the terms (a) and (b) above are both $o(1)$.

Part (a) Note that

$$\frac{1}{\epsilon^2 n^2}\sum_{i=1}^{n}\mathbb{E}\left[\tilde{f}_t(\mathcal{H}_t^{(i)})^2\right] = \frac{1}{\epsilon^2 n^2}\sum_{i=1}^{n}\left\{\mathbb{E}\left[W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})^2 f_t(\mathcal{H}_t^{(i)})^2\right] - \mathbb{E}_{\pi^*_{2:T}}\left[f_t(\mathcal{H}_t^{(i)})\right]^2\right\}$$

Since $W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})^2 \leq \pi_{\min}^{-(t-1)} W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})$ by Condition 1,

$$\leq \frac{\pi_{\min}^{-(t-1)}}{\epsilon^2 n^2}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)})^2\right] = \frac{\pi_{\min}^{-(t-1)}}{\epsilon^2 n^2}\sum_{i=1}^{n}\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})^2\right] \to 0.$$

The limit above holds because $\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})^2\right] < \infty$ by assumption.

Part (b) For any $s \in [1\colon t]$, we now show a helpful result for any function h of $\mathcal{H}_s^{(i)}$ and any constants $c^{(i)}, c^{(j)}$:

$$\mathbb{E}\Big[\Big\{W_{2:s}^{(i)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(i)}) - c^{(i)}\Big\}\Big\{W_{2:s}^{(j)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(j)}) - c^{(j)}\Big\}\Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[\Big\{W_{2:s}^{(i)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(i)}) - c^{(i)}\Big\}\Big\{W_{2:s}^{(j)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(j)}) - c^{(j)}\Big\}\Big|\mathcal{H}_{s-1}^{(1:n)}\Big]\Big]$$

Note that conditional on $\mathcal{H}_{s-1}^{(1:n)}$, the random part of $\mathcal{H}_s^{(i)}$ is $\{X_s^{(i)}, A_s^{(i)}, R_s^{(i)}\}$. Note that $\{X_s^{(i)}, A_s^{(i)}, R_s^{(i)}\}$ and $\{X_s^{(j)}, A_s^{(j)}, R_s^{(j)}\}$ are independent conditional on $\mathcal{H}_{s-1}^{(1:n)}$ for $i \neq j$.

$$= \mathbb{E}\Big[\mathbb{E}\Big[W_{2:s}^{(i)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(i)}) - c^{(i)}\Big|\mathcal{H}_{s-1}^{(1:n)}\Big]\mathbb{E}\Big[W_{2:s}^{(j)}(\theta^*_{1:s-1}, \hat{\theta}^{(n)}_{1:s-1})h(\mathcal{H}_s^{(j)}) - c^{(j)}\Big|\mathcal{H}_{s-1}^{(1:n)}\Big]\Big]$$

$$= \mathbb{E}\Big[\Big\{W_{2:s-1}^{(i)}(\theta^*_{1:s-2}, \hat{\theta}^{(n)}_{1:s-2})\mathbb{E}_{\pi_s^*}\Big[h(\mathcal{H}_s^{(i)})\Big|\mathcal{H}_{s-1}^{(1:n)}\Big] - c^{(i)}\Big\}$$
$$\Big\{W_{2:s-1}^{(j)}(\theta^*_{1:s-2}, \hat{\theta}^{(n)}_{1:s-2})\mathbb{E}_{\pi_s^*}\Big[h(\mathcal{H}_s^{(j)})\Big|\mathcal{H}_{s-1}^{(1:n)}\Big] - c^{(j)}\Big\}\Big]$$

$$= \mathbb{E}\Big[\Big\{W_{2:s-1}^{(i)}(\theta^*_{1:s-2}, \hat{\theta}^{(n)}_{1:s-2})\mathbb{E}_{\pi_s^*}\Big[h(\mathcal{H}_s^{(i)})\Big|\mathcal{H}_{s-1}^{(i)}\Big] - c^{(i)}\Big\}$$
$$\Big\{W_{2:s-1}^{(j)}(\theta^*_{1:s-2}, \hat{\theta}^{(n)}_{1:s-2})\mathbb{E}_{\pi_s^*}\Big[h(\mathcal{H}_s^{(j)})\Big|\mathcal{H}_{s-1}^{(j)}\Big] - c^{(j)}\Big\}\Big]. \quad (22)$$

We use Equation (22) above to simplify the following term:

$$\frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Big[\tilde{f}_t(\mathcal{H}_t^{(i)})\tilde{f}_t(\mathcal{H}_t^{(j)})\Big]$$

$$= \frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Big[\Big\{W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})f_t(\mathcal{H}_t^{(i)}) - \mathbb{E}_{\pi^*_{2:T}}\big[f_t(\mathcal{H}_t^{(i)})\big]\Big\}$$
$$\Big\{W_{2:t}^{(j)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})f_t(\mathcal{H}_t^{(j)}) - \mathbb{E}_{\pi^*_{2:t}}\big[f_t(\mathcal{H}_t^{(j)})\big]\Big\}\Big].$$

First, we apply Equation (22) for $h(\mathcal{H}_t^{(i)}) \triangleq f_t(\mathcal{H}_t^{(i)})$ and $c^{(i)} \triangleq \mathbb{E}_{\pi^*_{2:T}}\big[f_t(\mathcal{H}_t^{(i)})\big]$.

$$= \frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Big[\Big\{W_{2:t-1}^{(i)}(\theta^*_{1:t-2}, \hat{\theta}^{(n)}_{1:t-2})\mathbb{E}_{\pi_t^*}\big[f_t(\mathcal{H}_t^{(i)})\big|\mathcal{H}_{t-1}^{(i)}\big] - \mathbb{E}_{\pi^*_{2:T}}\big[f_t(\mathcal{H}_t^{(i)})\big]\Big\}$$
$$\Big\{W_{2:t-1}^{(j)}(\theta^*_{1:t-2}, \hat{\theta}^{(n)}_{1:t-2})\mathbb{E}_{\pi_t^*}\big[f_t(\mathcal{H}_t^{(j)})\big|\mathcal{H}_{t-1}^{(j)}\big] - \mathbb{E}_{\pi^*_{2:t}}\big[f_t(\mathcal{H}_t^{(j)})\big]\Big\}\Big]$$

By applying Equation (22) again for $h(\mathcal{H}_{t-1}^{(i)}) \triangleq \mathbb{E}_{\pi_t^*}\big[f_t(\mathcal{H}_t^{(i)})\big|\mathcal{H}_{t-1}^{(i)}\big]$ and $c^{(i)} \triangleq \mathbb{E}_{\pi_{2:T}^*}\big[f_t(\mathcal{H}_t^{(i)})\big]$,

$$= \frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Bigg[\Big\{W_{2:t-2}^{(i)}(\theta_{1:t-3}^*,\hat{\theta}_{1:t-3}^{(n)})\mathbb{E}_{\pi_t^*}\big[f_t(\mathcal{H}_t^{(i)})\big|\mathcal{H}_{t-2}^{(i)}\big] - \mathbb{E}_{\pi_{2:T}^*}\big[f_t(\mathcal{H}_t^{(i)})\big]\Big\}$$
$$\Big\{W_{2:t-2}^{(j)}(\theta_{1:t-3}^*,\hat{\theta}_{1:t-3}^{(n)})\mathbb{E}_{\pi_t^*}\big[f_t(\mathcal{H}_t^{(j)})\big|\mathcal{H}_{t-2}^{(j)}\big] - \mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(j)})\big]\Big\}\Bigg]$$

By repeatedly applying Equation (22), we have

$$= \frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Bigg[\Big\{\mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(i)})\big|\mathcal{H}_1^{(i)}\big] - \mathbb{E}_{\pi_{2:T}^*}\big[f_t(\mathcal{H}_t^{(i)})\big]\Big\}$$
$$\Big\{\mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(j)})\big|\mathcal{H}_1^{(j)}\big] - \mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(j)})\big]\Big\}\Bigg]$$

Since $\mathcal{H}_1^{(j)}$ and $\mathcal{H}_1^{(i)}$ are independent for $i \neq j$,

$$= \frac{2}{\epsilon^2 n^2}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathbb{E}\Big[\mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(i)})\big|\mathcal{H}_1^{(i)}\big] - \mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(i)})\big]\Big]$$
$$\mathbb{E}\Big[\mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(j)})\big|\mathcal{H}_1^{(j)}\big] - \mathbb{E}_{\pi_{2:t}^*}\big[f_t(\mathcal{H}_t^{(j)})\big]\Big] = 0.$$

The final equality above holds by law of iterated expectations. ■

Appendix C. Asymptotic Normality

Theorem 2 (Asymptotic Normality of Z-Estimator) *We consider the setting of Section 1. Recall that the result of Theorem 1 is that* $\hat{\theta}^{(n)} \xrightarrow{P} \theta^*$ *and* $\hat{\beta}_t^{(n)} \xrightarrow{P} \beta_t^*$ *for all* $t \in [1\colon T-1]$. *Moreover,* $\|\hat{\theta}^{(n)}(\,\cdot\,) - \theta^*(\,\cdot\,)\|_{B_{1:T-1}} \xrightarrow{P} 0$ *and* $\|\hat{\beta}_t^{(n)}(\,\cdot\,) - \beta_t^*(\,\cdot\,)\|_{B_{1:-1}} \xrightarrow{P} 0$ *for all* $t \in [1\colon T-1]$. *We prove that the following holds under Conditions 1, 2, 3, 5, 8, 9, 10 and the results of Theorem 1:*

$$\sqrt{n}\big(\hat{\theta}^{(n)} - \theta^*\big) \xrightarrow{D} \mathcal{N}\left(0, \dot{\Psi}^{-1} M^{\text{adaptive}} (\dot{\Psi}^{-1})^\top\right), \tag{8}$$

where $M^{\text{adaptive}} \triangleq \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\psi\big(\mathcal{H}_T^{(i)}; \theta^*\big) + \dot{\Psi} \sum_{t=1}^{T-1} \big(\frac{\partial \theta^*}{\partial \beta_t^*}\big) \dot{\Phi}_t^{-1} \phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\right\}^{\otimes 2}\right]$.

Note above our Theorem statement differs from that in the main text because we do not use Condition 4 and instead use Conditions 9 and 10; see Lemma 5 for a proof that Condition 4 implies that Conditions 9 and 10 hold.

Condition 1 (Minimum Exploration) *For some* $\pi_{\min} > 0$, *for all* $t \in [1\colon T]$,

$$\min_{a \in \mathcal{A}} \hat{\pi}_t\big(a, X_t^{(i)}\big) \geq \pi_{\min} \text{ w.p. } 1 \quad \text{and} \quad \inf_{\beta_{t-1} \in \mathbb{R}^{d_{t-1}}} \min_{a \in \mathcal{A}} \pi_t\big(a, X_t^{(i)}; \beta_{t-1}\big) \geq \pi_{\min} \text{ w.p. } 1.$$

Condition 2 (Locally Lipschitz Policy Function) *For all* $t \in [2\colon T]$, *there exists a function* $m_t\big(X_t^{(i)}\big)$ *such that (i)* $\mathbb{E}_{\pi^*_{2:t}}\big[m_t\big(X_t^{(i)}\big)\big] < \infty$ *and (ii) for all* $a \in \mathcal{A}$ *and for any* $\beta_t \in B_t$,

$$\Big|\pi_t\big(a, X_t^{(i)}; \beta_{t-1}\big) - \pi_t\big(a, X_t^{(i)}; \beta_{t-1}^*\big)\Big| \leq m_t\big(X_t^{(i)}\big)\big\|\beta_{t-1} - \beta_{t-1}^*\big\|.$$

Condition 3 (Finite Moments) *For some* $\alpha > 0$, *for all* $t \in [1\colon T]$,

$$\mathbb{E}_{\pi^*_{2:t}}\left[\big\|\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*\big)\big\|_1^{4+\alpha}\right] < \infty \quad \text{and} \quad \mathbb{E}_{\pi^*_{2:T}}\left[\big\|\psi_T\big(\mathcal{H}_T^{(i)}; \theta^*\big)\big\|_1^{4+\alpha}\right] < \infty.$$

Condition 4 (Lipschitz Estimating Functions) *There exists a function* $g_T\big(\mathcal{H}_T^{(i)}\big)$ *such that (i)* $\mathbb{E}_{\pi^*_{2:T}}\big[g_T\big(\mathcal{H}_T^{(i)}\big)^2\big] < \infty$ *and (ii) for all* $\beta_{1:T-1}, \beta'_{1:T-1} \in B_{1:T-1}$, $\theta, \theta' \in \Theta$,

$$\Big\|f_T\big(\mathcal{H}_T^{(i)}; \beta_{1:T-1}, \theta\big) - f_T\big(\mathcal{H}_T^{(i)}; \beta'_{1:T-1}, \theta'\big)\Big\| \leq g_T\big(\mathcal{H}_T^{(i)}\big)\,\big\|\big[\beta_{1:T-1}, \theta\big] - \big[\beta'_{1:T-1}, \theta'\big]\big\|\,.$$

Also for all $t \in [2\colon T]$, *there exists a function* $g_t\big(\mathcal{H}_t^{(i)}\big)$ *such that (i)* $\mathbb{E}_{\pi^*_{2:t}}\big[g_t\big(\mathcal{H}_t^{(i)}\big)^2\big] < \infty$ *and (ii) for all* $\beta_{1:t}, \beta'_{1:t} \in B_{1:t}$, $\|f_t(\mathcal{H}_t^{(i)}; \beta_{1:t}, \theta) - f_t(\mathcal{H}_t^{(i)}; \beta_{1:t})\| \leq g_t(\mathcal{H}_t^{(i)})\|\beta_{1:t} - \beta'_{1:t}\|$.

Condition 5 (Fréchet Differentiability) $\theta^{*,[s-1]}(\,\cdot\,)$ *is Fréchet differentiable with respect to* $\beta_s^*(\,\cdot\,)$ *and continuous in* $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ *for all* $s < T$, *and for all* $t \in [1\colon T-1]$, $\beta_t^{*,[s-1]}(\,\cdot\,)$ *is Fréchet differentiable with respect to* $\beta_s^*(\,\cdot\,)$ *and continuous in* $\beta_s(\,\cdot\,) : B_{1:s-1} \mapsto B_s$ *for all* $s < t$. *Also, the derivative functions* $\frac{\partial \theta^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ *and* $\frac{\partial \beta_t^{*,[s-1]}(\,\cdot\,)}{\partial \beta_s^*(\,\cdot\,)}$ *are continuous in their arguments* $\beta_{1:s-1} \in B_{1:s-1}$. $\mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ *is Fréchet differentiable with respect to* $\theta^*(\,\cdot\,)$, *and* $\mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ *is Fréchet differentiable with respect to* $\beta_t^*(\,\cdot\,)$ *for all* $t \in [1\colon T-1]$. *Also, derivative functions* $\frac{\partial}{\partial \theta^*(\,\cdot\,)} \mathbb{E}_{\pi_{2:T}(\,\cdot\,)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\,\cdot\,)\big)\big]$ *and* $\frac{\partial}{\partial \beta_t^*(\,\cdot\,)} \mathbb{E}_{\pi_{2:t}(\,\cdot\,)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\,\cdot\,)\big)\big]$ *are continuous in their arguments,* $\beta_{1:T-1} \in B_{1:T-1}$ *and* $\beta_{1:t-1} \in B_{1:t-1}$, *respectively.*

Condition 8 (Positive Definite Bread) *$\frac{\partial}{\partial \theta^*(\cdot)} \mathbb{E}_{\pi_{2:T}(\cdot)}\big[\psi\big(\mathcal{H}_T^{(i)}; \theta^*(\cdot)\big)\big]$ is finite and positive definite uniformly over $\beta_{1:T-1} \in B_{1:T-1}$. Also, $\frac{\partial}{\partial \beta_t^*(\cdot)} \mathbb{E}_{\pi_{2:t}(\cdot)}\big[\phi_t\big(\mathcal{H}_t^{(i)}; \beta_t^*(\cdot)\big)\big]$ is finite and positive definite uniformly over $\beta_{1:t-1} \in B_{1:t-1}$ for all $t \in [1\colon T-1]$.*

Recall that $f_T(\,\cdot\,; \beta_{1:T-1}, \theta) \triangleq \big(\prod_{t=2}^{T} \pi_t(\,\cdot\,; \beta_{t-1})\big)\psi(\,\cdot\,; \theta)$ and the functions $f_t(\,\cdot\,; \beta_{1:t}) \triangleq \big(\prod_{s=2}^{t-1} \pi_s(\,\cdot\,; \beta_{s-1})\big)\phi_t(\,\cdot\,; \beta_t)$. Also recall we defined $\mathcal{F}_{T,c_T} \triangleq \big\{c_T^\top f_T(\,\cdot\,; \beta_{1:T-1}, \theta) : \beta_{1:T-1} \in B_{T-1}, \theta \in \Theta\big\}$ for any fixed $c_T \in \mathbb{R}^{d_T}$. Similarly, we define $\mathcal{F}_{t,c_t} \triangleq \big\{c_t^\top f_t(\,\cdot\,; \beta_{1:t-1}, \theta) : \beta_{1:t-1} \in B_{t-1}\big\}$ for any fixed $c_t \in \mathbb{R}^{d_t}$.

Condition 9 (Finite Bracketing Integral) *For each $t \in [1\colon T]$, we assume that for any finite, fixed $c_t \in \mathbb{R}^{d_t}$,*

$$\int_0^1 \sqrt{\log N_{[\,]}\big(\epsilon, \mathcal{F}_{t,c_t}, L_2\big(\mathcal{P}_{\pi^*_{2:t}}\big)\big)}\, d\epsilon < \infty,$$

where $\mathcal{P}_{\pi^*_{2:t}}$ is the distribution of potential outcomes $\mathcal{P}$ where actions are selected with policies $\pi^*_{2:t}$. We also assume $\mathcal{F}_{t,c_t}$ has a measurable envelope function F_{t,c_t} with $\mathbb{E}_{\pi^*_{2:t}}\big[F_{t,c_t}(\mathcal{H}_t^{(i)})^2\big] < \infty$.

Above, recall that an envelope function F_{t,c_t} for the class of functions $\mathcal{F}_{t,c_t}$ means that $\sup_{f \in \mathcal{F}_{t,c_t}} |f(\mathcal{H}_t^{(i)})| < F_{t,c_t}(\mathcal{H}_t^{(i)}) < \infty$ w.p. 1 (Van der Vaart, 2000, pg. 270).

Below we use the semi-metric $\rho_t(f, f') \triangleq \mathbb{E}_{\pi^*_{2:t}}\big[\big\|f(\mathcal{H}_t^{(i)}) - f'(\mathcal{H}_t^{(i)})\big\|^2\big]$ for any $f, f' \in \mathcal{F}_{t,c_t}$.

Condition 10 (Locally Continuous Function Classes) *For any $\epsilon > 0$, there exists a $\delta_{T,\epsilon} > 0$ such that for $\beta_{1:T-1} \in \mathbb{R}^{\sum_{t=1}^{T-1} d_t}$ and $\theta \in \mathbb{R}^{d_T}$ with $\|[\beta_{1:T-1}, \theta] - [\beta^*_{1:T-1}, \theta^*]\| < \delta_{T,\epsilon}$,*

$$\rho_T\left(f_T(\,\cdot\,; \beta_{1:T-1}, \theta), f_T(\,\cdot\,; \beta^*_{1:T-1}, \theta^*)\right) < \epsilon.$$

Similarly, for all $t \in [1\colon T-1]$, for any $\epsilon > 0$, there exists a $\delta_{t,\epsilon} > 0$ such that for $\beta_{1:t} \in \mathbb{R}^{\sum_{s=1}^{t} d_s}$ with $\|\beta_{1:t} - \beta^*_{1:t}\| < \delta_{t,\epsilon}$, then $\rho_t\left(f_t(\,\cdot\,; \beta_{1:t}), f_t(\,\cdot\,; \beta^*_{1:t})\right) < \epsilon$.

Notation for the Remainder of this Section (Appendix C) For notational convenience, we let $\theta_T \triangleq \theta$ and $\theta_t \triangleq \beta_t$ for all $t \in [1\colon T-1]$. This means that $\theta_T^* \triangleq \theta^*$, $\theta_t^* \triangleq \beta_t^*$, $\hat{\theta}_T^{(n)} \triangleq \hat{\theta}^{(n)}$, and $\hat{\theta}_t^{(n)} \triangleq \hat{\beta}_t^{(n)}$. Also we use $\Theta_T \triangleq \Theta$ and $\Theta_t \triangleq B_t$, where recall Θ is a bounded ball that contains $\theta^*(\beta_{1:T-1})$ for all $\beta_{1:T-1} \in B_{1:T-1}$ and B_t is a bounded ball that contains $\beta_t^*(\beta_{1:t-1})$ for all $\beta_{1:t-1} \in B_{1:t-1}$. Additionally, we let $\psi_T \triangleq \psi$ and $\psi_t \triangleq \phi_t$ for all $t \in [1\colon T-1]$, so $\psi_T(\mathcal{H}_T^{(i)}; \theta_T) = \psi(\mathcal{H}_T^{(i)}; \theta)$ and $\psi_t(\mathcal{H}_t^{(i)}; \theta_t) = \phi_t(\mathcal{H}_t^{(i)}; \beta_t)$ for $t < T$. We also define the following notation:

$$W_{2:t}^{(i)}\big(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}\big) \triangleq \prod_{s=2}^{t} W_s^{(i)}\big(\theta_{s-1}, \hat{\theta}_{s-1}^{(n)}\big)$$

$$\Psi_t(\theta_{1:t}) \triangleq \mathbb{E}_{\pi_2(\theta_1), \pi_3(\theta_2), \ldots, \pi_t(\theta_{t-1})}\big[\psi_t(\mathcal{H}_t^{(i)}; \theta_t)\big] = \mathbb{E}\left[W_{2:t}^{(i)}\big(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}\big)\psi_t(\mathcal{H}_t^{(i)}; \theta_t)\right]$$

$$\hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big) \triangleq \frac{1}{n}\sum_{i=1}^{n} W_{2:t}^{(i)}\big(\theta_{1:t-1}, \hat{\theta}_{1:t-1}^{(n)}\big)\psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big)$$

C.1. Proof of Theorem 2

The main result we will show in the proof is Equation (23) below. We let $\dot{\Psi}_t \triangleq \mathbb{E}_{\pi^*_{2:T}}\left[\frac{\partial}{\partial\theta^*}\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big)\right]$ for $t \in [1\colon T]$. Consider the following:

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_1 - \theta_1^* \\ \hat{\theta}_2(\hat{\theta}_1) - \theta_2^*(\hat{\theta}_1) \\ \vdots \\ \hat{\theta}_{t-1}\big(\hat{\theta}_{1:t-2}\big) - \theta_{t-1}^*\big(\hat{\theta}_{1:t-2}\big) \\ \hat{\theta}_t\big(\hat{\theta}_{1:t-1}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}\big) \end{bmatrix} = -\sqrt{n}\begin{bmatrix} \dot{\Psi}_1^{-1}\hat{\Psi}_1^{(n)}(\theta_1^*) \\ \dot{\Psi}_2^{-1}\hat{\Psi}_2^{(n)}\big(\theta_{1:2}^*\big) \\ \vdots \\ \dot{\Psi}_{t-1}^{-1}\hat{\Psi}_{t-1}^{(n)}\big(\theta_{1:t-1}^*\big) \\ \dot{\Psi}_t^{-1}\hat{\Psi}_t^{(n)}\big(\theta_{1:t}^*\big) \end{bmatrix} + o_P(1)$$
$$\xrightarrow{D} \mathcal{N}\left(0, \left[\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top\right]_{u=1,s=1}^{u=t,s=t}\right), \quad (23)$$

where for $u, s \in [1\colon t]$, $\Sigma_{u,s} \triangleq \mathbb{E}_{\pi^*_{2:t}}\left[\psi_u\big(\mathcal{H}_u^{(i)};\theta_u^*\big)\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)^\top\right] \in \mathbb{R}^{d_u \times d_s}$, and we use the notation $[V_{u,s}]_{u=1,s=1}^{u=t,s=t} \triangleq \begin{bmatrix} V_{1,1} & V_{1,2} & \dots & V_{1,t} \\ V_{2,1} & V_{2,2} & \dots & V_{2,t} \\ \vdots & \vdots & \ddots & \vdots \\ V_{t,1} & V_{t,2} & \dots & V_{t,t} \end{bmatrix}$ for matrices $V_{u,s}$.

Lemma 9 (see Section C.3) states that under Conditions 1, 2, 3, 5, 8, 9, 10, and the condition that $\sqrt{n}\big(\hat{\theta}_s^{(n)} - \theta_s^*\big) = O_P(1)$ for all $s < t$, that Equation (23) above implies that

$$\sqrt{n}\left(\hat{\theta}_t\big(\hat{\theta}_{1:t-1}\big) - \theta_t^*\big(\theta_{1:t-1}^*\big)\right) \xrightarrow{D} \mathcal{N}\left(0, \sum_{u=1}^{t}\sum_{s=1}^{t}\frac{\partial\theta_t^*}{\partial\theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top\frac{\partial\theta_t^*}{\partial\theta_s^*}^\top\right).$$

We now show that $\sum_{u=1}^{t}\sum_{s=1}^{t}\frac{\partial\theta_t^*}{\partial\theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top\frac{\partial\theta_t^*}{\partial\theta_s^*}^\top = \dot{\Psi}_t^{-1}M_t^{\text{adaptive}}(\dot{\Psi}_t^{-1})^\top$, where recall that $M_t^{\text{adaptive}} \triangleq \mathbb{E}_{\pi^*_{2:t}}\left[\left\{\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big) + \dot{\Psi}_t\sum_{s=1}^{t-1}\big(\frac{\partial\theta_t^*}{\partial\theta_s^*}\big)\dot{\Psi}_s^{-1}\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)\right\}^{\otimes 2}\right]$.

$$M_t^{\text{adaptive}} = \mathbb{E}_{\pi^*_{2:t}}\left[\left\{\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big) + \dot{\Psi}_t\sum_{s=1}^{t-1}\left(\frac{\partial\theta_t^*}{\partial\theta_s^*}\right)\dot{\Psi}_s^{-1}\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)\right\}^{\otimes 2}\right]$$

$$= \mathbb{E}_{\pi^*_{2:t}}\left[\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big)^{\otimes 2}\right] + \mathbb{E}_{\pi^*_{2:t}}\left[\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big)\left\{\sum_{s=1}^{t-1}\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)^\top(\dot{\Psi}_s^{-1})^\top\left(\frac{\partial\theta_t^*}{\partial\theta_s^*}\right)^\top\dot{\Psi}_t^\top\right\}\right]$$
$$+ \mathbb{E}_{\pi^*_{2:t}}\left[\left\{\dot{\Psi}_t\sum_{s=1}^{t-1}\left(\frac{\partial\theta_t^*}{\partial\theta_s^*}\right)\dot{\Psi}_s^{-1}\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)\right\}\psi_t\big(\mathcal{H}_t^{(i)};\theta_t^*\big)^\top\right]$$
$$+ \dot{\Psi}_t\mathbb{E}_{\pi^*_{2:t}}\left[\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\left(\frac{\partial\theta_t^*}{\partial\theta_s^*}\right)\dot{\Psi}_s^{-1}\psi_s\big(\mathcal{H}_s^{(i)};\theta_s^*\big)\psi_{s'}\big(\mathcal{H}_{s'}^{(i)};\theta_{s'}^*\big)^\top(\dot{\Psi}_{s'}^{-1})^\top\left(\frac{\partial\theta_t^*}{\partial\theta_s^*}\right)^\top\right]\dot{\Psi}_t^\top$$

$$= \Sigma_{t,t} + \sum_{s=1}^{t-1} \Sigma_{t,s} (\dot{\Psi}_s^{-1})^\top \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right)^\top \dot{\Psi}_t^\top + \dot{\Psi}_t \sum_{s=1}^{t-1} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right) \dot{\Psi}_s^{-1} \Sigma_{t,s} + \dot{\Psi}_t \left(\sum_{s=1}^{t-1} \sum_{s'=1}^{t-1} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right) \dot{\Psi}_s^{-1} \Sigma_{s,s'} (\dot{\Psi}_{s'}^{-1})^\top \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right)^\top \right) \dot{\Psi}_t^\top$$

By the above, we have that $\dot{\Psi}_t^{-1} M_t^{\text{adaptive}} \dot{\Psi}_t^{-1}$ equals the following

$$\dot{\Psi}_t^{-1} \Sigma_{t,t} (\dot{\Psi}_t^{-1})^\top + \dot{\Psi}_t^{-1} \sum_{s=1}^{t-1} \Sigma_{t,s} \dot{\Psi}_s^{-1} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right)^\top + \sum_{s=1}^{t-1} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right) \dot{\Psi}_s^{-1} \Sigma_{t,s} (\dot{\Psi}_t^{-1})^\top + \sum_{s=1}^{t-1} \sum_{s'=1}^{t-1} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right) \dot{\Psi}_s^{-1} \Sigma_{s,s'} (\dot{\Psi}_{s'}^{-1})^\top \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right)^\top$$

Since $\frac{\partial \theta_t^*}{\partial \theta_t^*} = I_{d_t}$ (identity function),

$$= \sum_{s=1}^{t} \sum_{s'=1}^{t} \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right) \dot{\Psi}_s^{-1} \Sigma_{s,s'} (\dot{\Psi}_{s'}^{-1})^\top \left(\frac{\partial \theta_t^*}{\partial \theta_s^*}\right)^\top.$$

Thus, we have that $\sqrt{n}\left(\hat{\theta}_t(\hat{\theta}_{1:t-1}) - \theta_t^*(\theta_{1:t-1}^*)\right) \xrightarrow{D} \mathcal{N}\left(0, \dot{\Psi}_t^{-1} M_t^{\text{adaptive}} (\dot{\Psi}_t^{-1})^\top\right)$. Equation (23) for $t = T$ is sufficient for the theorem, i.e., Equation (8). We will prove that Equation (23) holds for arbitrary t using an induction argument. Specifically we will first show the result holds for $t = 1$. Then we will show the result holds for arbitrary t, assuming the result already holds for all values of less than t, i.e., for $t-1, t-2, ..., 1$.

C.1.1. BASE CASE

By consistency of $\hat{\theta}_1$, finite second moment of $\psi(\mathcal{H}_1^{(i)}; \theta_1^*)$ (condition 3), invertibility of $\dot{\Psi}_1$ (condition 5), and finite bracketing integral for $\mathcal{F}_1$ (condition 9), by Van der Vaart (2000, Theorem 19.5), an asymptotic normality result for Z-Estimators in the i.i.d. case,

$$\sqrt{n}(\hat{\theta}_1 - \theta_1^*) = -\dot{\Psi}_1^{-1} \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_1(\mathcal{H}_1^{(i)}; \theta_1^*)}_{\sqrt{n}\hat{\Psi}_1^{(n)}(\theta_1^*)} + o_P(1) \xrightarrow{D} \mathcal{N}\left(0, \dot{\Psi}_1^{-1} \Sigma_{1,1} (\dot{\Psi}_t^{-1})^\top\right). \tag{24}$$

Note that $\dot{\Psi}_1^{-1} \Sigma_{1,1} (\dot{\Psi}_1^{-1})^\top = \frac{\partial \theta_1^*}{\partial \theta_1^*} \dot{\Psi}_1^{-1} \Sigma_{1,1} (\dot{\Psi}_1^{-1})^\top \frac{\partial \theta_1^*}{\partial \theta_1^*}^\top$. Thus, we have that Equation (23) holds for the $t = 1$ case.

C.1.2. INDUCTION STEP

We now assume that Equation (23) holds for the $t-1, t-2, \ldots 1$ cases and will show that it holds for the t^{th} case.

Fix vector $c_t \in \mathbb{R}^{d_t}$. Consider the following stochastic process (recall that the class of functions $\mathcal{F}_{t,c_t}$ is defined above Condition 9):

$$\left\{\sqrt{n}c_t^\top\left[\hat{\Psi}_t^{(n)}(\theta_{1:t}) - \Psi_t(\theta_{1:t})\right] : \theta_t \in \Theta_t \text{ for all } \in [1\colon t]\right\}$$
$$= \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\prod_{s=2}^{t}\hat{\pi}_s(A_s^{(i)}, X_s^{(i)})\right)^{-1} f(\mathcal{H}_t^{(i)}) - \mathbb{E}\left[\left(\prod_{s=2}^{t}\hat{\pi}_s(A_s^{(i)}, X_s^{(i)})\right)^{-1} f(\mathcal{H}_t^{(i)})\right] : f \in \mathcal{F}_{t,c_t}\right\}.$$

Now we apply one of our most critical results, Lemma 16, which states that under under Conditions 1, 2, 3, 9, 10, when $\|\hat{\theta}_t^{(n)}(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{\Theta_{1:t-1}} \xrightarrow{P} 0$ and $\hat{\theta}_s^{(n)} \xrightarrow{P} \theta_s^*$ for all $s \in [1\colon t-1]$, then for any fixed $c_t \in \mathbb{R}^{d_t}$,

$$\sqrt{n}\Bigg\|c_t^\top\left[\hat{\Psi}_t^{(n)}(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,)) - \Psi_t(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,))\right] - c_t^\top\left[\hat{\Psi}_t^{(n)}(\,\cdot\,, \theta_t^*(\,\cdot\,)) - \Psi_t(\,\cdot\,, \theta_t^*(\,\cdot\,))\right]\Bigg\|_{\Theta_{1:t-1}} \xrightarrow{P} 0.$$

Note that the conditions of Lemma 16 hold by the assumptions of this Theorem. Since $\mathbb{P}(\hat{\theta}_{1:t}^{(n)} \in \Theta_{1:t}) \to 1$ by consistency of $\hat{\theta}_{1:t}^{(n)}$ (an assumption of this Theorem), the result of the Lemma implies that

$$\sqrt{n}c_t^\top\left[\hat{\Psi}_t^{(n)}(\hat{\theta}_{1:t}^{(n)}) - \Psi_t(\hat{\theta}_{1:t}^{(n)})\right] = \sqrt{n}c_t^\top\left[\hat{\Psi}_t^{(n)}(\theta_{1:t}^*) - \Psi_t(\theta_{1:t}^*)\right] + o_P(1). \tag{25}$$

Again, since $\mathbb{P}(\hat{\theta}_{1:t}^{(n)} \in \Theta_{1:t}) \to 1$, by Fréchet Differentiability Condition 5,

$$\sqrt{n}\left\{\Psi_t(\hat{\theta}_{1:t}^{(n)}) - \Psi_t\left(\hat{\theta}_{1:t-1}^{(n)}, \theta_t^*(\hat{\theta}_{1:t-1}^{(n)})\right)\right\}$$
$$= \frac{\partial}{\partial\theta_t^*(\hat{\theta}_{1:t-1}^{(n)})}\Psi_t\left(\hat{\theta}_{1:t-1}^{(n)}, \theta_t^*(\hat{\theta}_{1:t-1}^{(n)})\right)\sqrt{n}\left(\hat{\theta}_t^{(n)} - \theta_t^*(\hat{\theta}_{1:t-1}^{(n)})\right)$$
$$+ \sqrt{n}o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}}\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\|\right) + o_P(1). \tag{26}$$

Note the following observations:

- Note by Condition 5, $\frac{\partial}{\partial\theta_t^*(\theta_{1:t-1})}\Psi_t\left(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\right)$ is continuous in $\theta_{1:t-1} \in \Theta_{1:t-1}$. Since $\hat{\theta}_{1:t}^{(n)} \xrightarrow{P} \theta_{1:t}^*$ (which holds by Theorem 1), by continuous mapping theorem,
$$\frac{\partial}{\partial\theta_t^*(\hat{\theta}_{1:t-1}^{(n)})}\Psi_t\left(\hat{\theta}_{1:t-1}^{(n)}, \theta_t^*(\hat{\theta}_{1:t-1}^{(n)})\right) \xrightarrow{P} \frac{\partial}{\partial\theta_t^*(\theta_{1:t-1}^*)}\Psi_t\left(\theta_{1:t-1}^*, \theta_t^*(\theta_{1:t-1}^*)\right) = \dot{\Psi}_t$$
- $\Psi_t\left(\hat{\theta}_{1:t-1}^{(n)}, \theta_t^*(\hat{\theta}_{1:t-1}^{(n)})\right) = 0$ and $\hat{\Psi}_t^{(n)}(\hat{\theta}_{1:t}^{(n)}) = 0$ by the definitions of $\theta_t^*(\,\cdot\,)$ and $\hat{\theta}_t^{(n)}(\,\cdot\,)$.
- Lemma 8, which states that $\sqrt{n}o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}}\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\|\right) = o_P(1)$ under the results of Theorem 1, Conditions 1, 2, 3, 5, 8, 9, 10, and the condition that $\sqrt{n}(\hat{\theta}_s^{(n)} - \theta_s^*) = O_P(1)$ for all $s \in [1\colon t-1]$. Note that these conditions are satisfied because $\sqrt{n}(\hat{\theta}_s^{(n)} - \theta_s^*) = O_P(1)$ for all $s \in [1\colon t-1]$ by our induction assumption by the argument below Equation (23).

By the above observations, Equation (26) implies that

$$-\sqrt{n}c_t^\top \left\{\hat{\Psi}_t^{(n)}\big(\hat{\theta}_{1:t}^{(n)}\big) - \Psi_t\big(\hat{\theta}_{1:t}^{(n)}\big)\right\} = c_t^\top \dot{\Psi}_t \sqrt{n}\left(\hat{\theta}_t^{(n)} - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big)\right) + o_P(1). \tag{27}$$

By Equations (27) and (25),

$$-\sqrt{n}c_t^\top \big[\hat{\Psi}_t^{(n)}\left(\theta_{1:t}^*\right) - \underbrace{\Psi_t\big(\theta_{1:t}^*\big)}_{=0}\big] = c_t^\top \dot{\Psi}_t \sqrt{n}\left(\hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big)\right) + o_P(1). \tag{28}$$

Note that by our induction assumption that Equation (23) holds for the $t-1, t-2, \ldots 1$ cases, we have that for each $s < t$,

$$\sqrt{n}\left(\hat{\theta}_s^{(n)}\big(\hat{\theta}_{1:s-1}^{(n)}\big) - \theta_s^*\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right) = -\sqrt{n}\dot{\Psi}_s^{-1}\hat{\Psi}_s^{(n)}\big(\theta_{1:s}^*\big) + o_P(1).$$

Pick any fixed vectors $c_s \in \mathbb{R}^{d_s}$ for $s \in [1\colon t-1]$. By multiplying by $c_s^\top \dot{\Psi}_s$ and summing over $s \in [1\colon t-1]$,

$$\sqrt{n}\sum_{s=1}^{t-1} c_s^\top \dot{\Psi}_s \left(\hat{\theta}_s^{(n)}\big(\hat{\theta}_{1:s-1}^{(n)}\big) - \theta_s^*\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right) = -\sqrt{n}\sum_{s=1}^{t-1} c_s^\top \hat{\Psi}_s^{(n)}\big(\theta_{1:s}^*\big) + o_P(1). \tag{29}$$

By summing Equations (28) and (29),

$$\sqrt{n}\sum_{s=1}^{t} c_s^\top \dot{\Psi}_s \left(\hat{\theta}_s^{(n)}\big(\hat{\theta}_{1:s-1}^{(n)}\big) - \theta_s^*\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right) + o_P(1) = -\sqrt{n}\sum_{s=1}^{t} c_s^\top \hat{\Psi}_s^{(n)}\big(\theta_{1:s}^*\big).$$

We now apply Importance-Weighted Martingale Central Limit Theorem (Lemma 10) to show that the above is asymptotically normal. Specifically, under Conditions 1, 2, 3 and the condition that $\sqrt{n}\big(\hat{\theta}_s^{(n)} - \theta_s^*\big) = O_P(1)$ for all $s \in [1\colon t-1]$ (this holds by our induction assumption; see the argument below Equation Equation (23)), Lemma 10 implies the following holds:

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{c_1^\top \psi_1\big(\mathcal{H}_1^{(i)}; \theta_1^*\big) + \sum_{s=2}^{t} W_{2:s}^{(i)}\big(\theta_{1:s-1}^*, \hat{\theta}_{1:s-1}^{(n)}\big) c_s^\top \psi_s\big(\mathcal{H}_s^{(i)}; \theta_s^*\big)\right\}$$
$$= -\sqrt{n}\sum_{s=1}^{t} c_s^\top \hat{\Psi}_s^{(n)}\big(\theta_{1:s}^*\big) \xrightarrow{D} \mathcal{N}\left(0, \sum_{u=1}^{t}\sum_{s=1}^{t} c_u^\top \, \Sigma_{u,s} \, c_s\right).$$

By Cramer Wold device,

$$\sqrt{n}\begin{bmatrix} \dot{\Psi}_1\big(\hat{\theta}_1^{(n)} - \theta_1^*\big) \\ \dot{\Psi}_2\left(\hat{\theta}_2^{(n)}\big(\hat{\theta}_1^{(n)}\big) - \theta_2^*\big(\hat{\theta}_1^{(n)}\big)\right) \\ \vdots \\ \dot{\Psi}_{t-1}\left(\hat{\theta}_{t-1}^{(n)}\big(\hat{\theta}_{1:t-2}^{(n)}\big) - \theta_{t-1}^*\big(\hat{\theta}_{1:t-2}^{(n)}\big)\right) \\ \dot{\Psi}_t\left(\hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big)\right) \end{bmatrix} + o_P(1) = -\sqrt{n}\begin{bmatrix} \hat{\Psi}_1^{(n)}\big(\theta_1^*\big) \\ \hat{\Psi}_2^{(n)}\big(\theta_{1:2}^*\big) \\ \vdots \\ \hat{\Psi}_{t-1}^{(n)}\big(\theta_{1:t-1}^*\big) \\ \hat{\Psi}_t^{(n)}\big(\theta_{1:t}^*\big) \end{bmatrix}$$
$$\xrightarrow{D} \mathcal{N}\left(0, [\Sigma_{u,s}]_{u=1,s=1}^{u=t,s=t}\right). \tag{30}$$

Since $\dot{\Psi}_s$ is invertible by Condition 8,

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_1^{(n)} - \theta_1^* \\ \hat{\theta}_2^{(n)}(\hat{\theta}_1^{(n)}) - \theta_2^*(\hat{\theta}_1^{(n)}) \\ \vdots \\ \hat{\theta}_{t-1}^{(n)}(\hat{\theta}_{1:t-2}^{(n)}) - \theta_{t-1}^*(\hat{\theta}_{1:t-2}^{(n)}) \\ \hat{\theta}_t^{(n)}(\hat{\theta}_{1:t-1}^{(n)}) - \theta_t^*(\hat{\theta}_{1:t-1}^{(n)}) \end{bmatrix} + o_P(1) \xrightarrow{D} \mathcal{N}\left(0, \left[\dot{\Psi}_u^{-1}\Sigma_{u,s}\dot{\Psi}_s^{-1}\right]_{u=1,s=1}^{u=t,s=t}\right). \tag{31}$$

Thus, by Equation (31) and Slutsky's Theorem, Equation (23) holds for the t^{th} case given the $t-1, t-2, \ldots, 1$ cases (induction step). ■

C.2. Lemma 8: $\sqrt{n}o_P(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\|)$ Converges to Zero

Lemma 8 ($\sqrt{n}o_P(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\|)$ Converges to Zero) *We consider the setting of Section 1. Recall that the result of Theorem 1 is that $\hat{\theta}_t^{(n)} \xrightarrow{P} \theta_t^*$ for all $t \in [1\colon T]$. Moreover, $\|\hat{\theta}_t^{(n)}(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{B_{1:T-1}} \xrightarrow{P} 0$ for all $t \in [1\colon T-1]$. Assuming the results of Theorem 1, Conditions 1, 2, 3, 5, 8, 9, 10, and that $\sqrt{n}(\hat{\theta}_s^{(n)} - \theta_s^*) = O_P(1)$ for all $s \in [1\colon t-1]$,*

$$\sqrt{n}o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right\|\right) = o_P(1).$$

Proof of Lemma 8: First note the following helpful result we will use. Since $\Psi_t\left(\theta_{1:t-2}, \theta_t^*(\theta_{1:t-2})\right)$ is Frechet differentiable with respect to $\theta_t^*(\theta_{1:t-2})$, by Condition 5,

$$\begin{aligned} \sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\Big\| &\Psi_t\left(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\right) - \Psi_t\left(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\right) \\ &\frac{\partial}{\partial\theta_t^*(\theta_{1:t-2})}\Psi_t\left(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\right)\sqrt{n}\left(\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right)\Big\| \\ = \sqrt{n}o_P&\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right\|\right). \end{aligned} \tag{32}$$

Now for the main argument.

$$\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right\|$$

Note that by Condition 8, $\frac{\partial}{\partial\theta_t^*(\theta_{1:t-1})}\Psi_t\left(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\right)$ exists and is positive definite uniformly over $\theta_{1:t-1} \in \Theta_{1:t-1}$, so for some $\lambda_{\min} > 0$,

$$\leq \lambda_{\min}^{-1} \sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\|\frac{\partial}{\partial\theta_t^*(\theta_{1:t-1})}\Psi_t\left(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\right)\sqrt{n}\left(\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right)\right\|$$

By Equation (32) and triangle inequality,

$$\leq \lambda_{\min}^{-1} \sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \Psi_t\big(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) \right\| + \lambda_{\min}^{-1}\sqrt{n} o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \right)$$

Since $\hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) = 0$ and $\Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) = 0$ by definitions of $\hat{\theta}_t^{(n)}(\,\cdot\,), \theta_t^*(\,\cdot\,)$,

$$\leq \lambda_{\min}^{-1} \sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\| \hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) \right\| + \lambda_{\min}^{-1}\sqrt{n} o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \right) \quad (33)$$

By the results of Theorem 1, $\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \xrightarrow{P} 0$. Lemma 16 states that under Conditions 1, 2, 3, 9, 10 and the conditions that $\hat{\theta}_s^{(n)} \xrightarrow{P} \theta_s^*$ for all $s \in [1\colon t-1]$ and $\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\| \xrightarrow{P} 0$, then for any fixed $c_t \in \mathbb{R}^{d_t}$,

$$\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}\hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \hat{\theta}_t^{(n)}(\theta_{1:t-1})\big) \right] - \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) \right] \right\| \xrightarrow{P} 0.$$

Note that the conditions of Lemma 16 are satisfied by the assumptions of this Lemma. Thus we have that by triangle inequality, that we can upper bound Equation (33) as follows:

$$\lambda_{\min}^{-1} \sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\| \hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) \right\| + o_P(1) + \lambda_{\min}^{-1}\sqrt{n} o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \right)$$

We apply now Theorem 15, which by Conditions 1, 2, 3, 9 and the condition that $\sqrt{n}(\hat{\theta}_s^{(n)} - \theta_s^*) = O_P(1)$ for all $s \in [1\colon t-1]$, implies that the following stochastic process is functionally asymptotically normal: $\left\{ \sqrt{n}\left[\hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) \right] : \theta_s \in \Theta_s \text{ for all } s \in [1\colon t] \right\}$. Thus, we have that, $\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\| \hat{\Psi}_t^{(n)}\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) - \Psi_t\big(\theta_{1:t-1}, \theta_t^*(\theta_{1:t-1})\big) \right\| = O_P(1)$. So,

$$= O_P(1) + o_P(1) + \lambda_{\min}^{-1}\sqrt{n} o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \right)$$

Summarizing the above results:

$$\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \leq O_P(1) + O(1)\sqrt{n} o_P\left(\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \left\| \hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1}) \right\| \right).$$

The above implies that $\sup_{\theta_{1:t-1}\in\Theta_{1:t-1}} \sqrt{n}\left\|\hat{\theta}_t^{(n)}(\theta_{1:t-1}) - \theta_t^*(\theta_{1:t-1})\right\| = O_P(1)$, which implies our desired result. ■

C.3. Lemma 9: Complicated Application of Delta Method

Lemma 9 (Complicated Application of Delta Method) *Recall that the result of Theorem 1 is that $\hat{\theta}_t^{(n)} \xrightarrow{P} \theta_t^*$ for all $t \in [1\colon T]$. Moreover, $\|\hat{\theta}_t^{(n)}(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{B_{1:T-1}} \xrightarrow{P} 0$ for all $t \in [1\colon T-1]$. Assuming the results of Theorem 1, Conditions 1, 2, 3, 5, 8, 9, 10, and the condition that $\sqrt{n}\big(\hat{\theta}_s^{(n)} - \theta_s^*\big) = O_P(1)$ for all $s \in [1\colon t-1]$ and*

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_1^{(n)} - \theta_1^* \\ \hat{\theta}_2^{(n)}(\hat{\theta}_1^{(n)}) - \theta_2^*(\hat{\theta}_1^{(n)}) \\ \vdots \\ \hat{\theta}_{t-1}^{(n)}\big(\hat{\theta}_{1:t-2}^{(n)}\big) - \theta_{t-1}^*\big(\hat{\theta}_{1:t-2}^{(n)}\big) \\ \hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big) \end{bmatrix} \xrightarrow{D} \mathcal{N}\left(0, \left[\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top\right]_{u=1,s=1}^{u=t,s=t}\right), \tag{34}$$

for some $t \in [1\colon T]$, then

$$\sqrt{n}\left(\hat{\theta}_t^{(n)} - \theta_t^*\right) \xrightarrow{D} \mathcal{N}\left(0, \sum_{u=1}^{t}\sum_{s=1}^{t} \frac{\partial\theta_t^*}{\partial\theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top \frac{\partial\theta_t^*}{\partial\theta_s^*}^\top\right). \tag{35}$$

Proof of Lemma 9: We will use the functions with superscripts defined before Condition 5 in the main text. Recall superscripts correspond to the number of arguments the function takes in. We will show that for any $s \in [1\colon t]$ that

$$\sqrt{n}\left(\theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big) - \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right) = \frac{\partial\theta_t^*}{\partial\theta_s^*}\sqrt{n}\left(\hat{\theta}_s^{(n)}(\hat{\theta}_{1:s-1}^{(n)}) - \theta_s^*(\hat{\theta}_{1:s-1}^{(n)})\right) + o_P(1). \tag{36}$$

We now show that Equation (36) above is sufficient for our desired result Equation (35). Equations (36) and (34) with Slutsky's theorem imply that

$$\sqrt{n}\begin{bmatrix} \theta_t^{*,[1]}\big(\hat{\theta}_1^{(n)}\big) - \theta_t^* \\ \theta_t^{*,[2]}\big(\hat{\theta}_{1:2}^{(n)}\big) - \theta_t^{*,[1]}\big(\hat{\theta}_1^{(n)}\big) \\ \theta_t^{*,[3]}\big(\hat{\theta}_{1:3}^{(n)}\big) - \theta_t^{*,[2]}\big(\hat{\theta}_{1:2}^{(n)}\big) \\ \vdots \\ \theta_t^{*,[t-1]}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^{*,[t-2]}\big(\hat{\theta}_{1:t-2}^{(n)}\big) \\ \hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big) \end{bmatrix} \xrightarrow{D} \mathcal{N}\left(0, \left[\frac{\partial\theta_t^*}{\partial\theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top \frac{\partial\theta_t^*}{\partial\theta_s^*}^\top\right]_{u=1,s=1}^{u=t,s=t}\right). \tag{37}$$

Note that $\frac{\partial\theta_t^*}{\partial\theta_t^*} = I_{d_t}$ (identity function). By summing the terms on the left hand side of Equation (37), we have the following telescoping series:

$$\begin{aligned}\sqrt{n}\left[\hat{\theta}_t^{(n)} - \theta_t^*\right] &= \sqrt{n}\left[\hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\theta_{1:t-1}^*\big)\right] \\ &= \sqrt{n}\left[\hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big) - \theta_t^*\big(\hat{\theta}_{1:t-1}^{(n)}\big)\right] + \sum_{s=1}^{t-1}\sqrt{n}\left[\theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big) - \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right]\end{aligned}$$

Thus, Equation (35) follows from Equation (36). We now show that Equation (36) holds. Recall that we showed the $t = 2$ case in the main text. To make the result clear, here we first show the result holds for the $t = 3$ case. We then prove the result holds for the general t case using an induction argument.

Showing Equation (36) Holds for $t = 3$ Case: We assume the $t = 2$ case and show that the $t = 3$ case holds, i.e., we assume the following

$$\sqrt{n}\left(\hat{\theta}_2^{(n)} - \theta_2^*\right) \xrightarrow{D} \mathcal{N}\left(0, \sum_{u=1}^{2}\sum_{s=1}^{2} \frac{\partial \theta_t^*}{\partial \theta_u^*} \dot{\Psi}_u^{-1} \Sigma_{u,s} (\dot{\Psi}_s^{-1})^\top \frac{\partial \theta_t^*}{\partial \theta_s^*}^\top\right). \tag{38}$$

and show that the following holds:

$$\sqrt{n}\left(\hat{\theta}_3^{(n)} - \theta_3^*\right) \xrightarrow{D} \mathcal{N}\left(0, \sum_{u=1}^{3}\sum_{s=1}^{3} \frac{\partial \theta_t^*}{\partial \theta_u^*} \dot{\Psi}_u^{-1} \Sigma_{u,s} (\dot{\Psi}_s^{-1})^\top \frac{\partial \theta_t^*}{\partial \theta_s^*}^\top\right). \tag{39}$$

Note that by telescoping series,

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_3^{(n)} - \theta_3^*\right) &= \sqrt{n}\left[\hat{\theta}_3^{(n)}(\hat{\theta}_{1:2}^{(n)}) - \theta_3^*(\theta_{1:2}^*)\right] \\ = \sqrt{n}\underbrace{\left[\hat{\theta}_3^{(n)}(\hat{\theta}_{1:2}^{(n)}) - \theta_3^*(\hat{\theta}_{1:2}^{(n)})\right]}_{\hat{\theta}_3 \text{ vs. } \theta_3^*} &+ \sqrt{n}\underbrace{\left[\theta_3^*(\hat{\theta}_{1:2}^{(n)}) - \theta_3^{*,[1]}(\hat{\theta}_1^{(n)})\right]}_{\hat{\theta}_2 \text{ vs. } \theta_2^*} + \sqrt{n}\underbrace{\left[\theta_3^{*,[1]}(\hat{\theta}_1^{(n)}) - \theta_3^*(\theta_{1:2}^*)\right]}_{\hat{\theta}_1 \text{ vs. } \theta_1^*}.\end{aligned} \tag{40}$$

Also note that by assumption of the Lemma, we have that

$$\sqrt{n}\begin{bmatrix} \hat{\theta}_1^{(n)} - \theta_1^* \\ \hat{\theta}_2^{(n)}(\hat{\theta}_1^{(n)}) - \theta_2^*(\hat{\theta}_1^{(n)}) \\ \hat{\theta}_3^{(n)}(\hat{\theta}_{1:2}^{(n)}) - \theta_3^*(\hat{\theta}_{1:2}^{(n)}) \end{bmatrix} \xrightarrow{D} \mathcal{N}\left(0, \left[\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top\right]_{u=1,s=1}^{u=3,s=3}\right). \tag{41}$$

Consider the following:

$$\sqrt{n}\begin{bmatrix} \theta_3^{*,[1]}(\hat{\theta}_1^{(n)}) - \theta_3^*(\theta_{1:2}^*) \\ \theta_3^*(\hat{\theta}_{1:2}^{(n)}) - \theta_3^{*,[1]}(\hat{\theta}_1^{(n)}) \\ \hat{\theta}_3^{(n)}(\hat{\theta}_{1:2}^{(n)}) - \theta_3^*(\hat{\theta}_{1:2}^{(n)}) \end{bmatrix} = \sqrt{n}\begin{bmatrix} \frac{\partial \theta_3^*}{\partial \theta_1^*}\left[\hat{\theta}_1^{(n)} - \theta_1^*\right] \\ \frac{\partial \theta_3^*}{\partial \theta_2^*}\left[\hat{\theta}_2^{(n)}(\hat{\theta}_1^{(n)}) - \theta_2^*(\hat{\theta}_1^{(n)})\right] \\ \hat{\theta}_3^{(n)}(\hat{\theta}_{1:2}^{(n)}) - \theta_3^*(\hat{\theta}_{1:2}^{(n)}) \end{bmatrix} + o_P(1). \tag{42}$$

Note we can derive the asymptotic distribution of the above using Equation (41). Additionally, note that by summing the terms on the left hand side above we get the telescoping series from Equation (40). Thus, to show Equation (39) holds, it is sufficient to show the above result.

Note the function $\theta_3^{*,[1]}(\cdot) = \theta_3^*(\cdot, \theta_2^*) : \mathbb{R}^{d_1} \mapsto \mathbb{R}^{d_3}$ where for any $\theta_1 \in \mathbb{R}^{d_1}$ we have that $\theta_3^{*,[1]}(\theta_1) = \theta_3^*(\theta_1, \theta_2^*) \triangleq \theta_3^*(\theta_1, \theta_2^*(\theta_1))$. By Condition 5, we have that $\frac{\partial \theta_3^*}{\partial \theta_1^*} \triangleq \frac{\partial \theta_3^*(\theta_1^*, \theta_2^*(\theta_1^*))}{\partial \theta_1^*}$ exists so by the Delta method (Van der Vaart, 2000, Theorem 3.1) we have that

$$\sqrt{n}\left[\theta_3^{*,[1]}(\hat{\theta}_1^{(n)}) - \theta_3^*\right] = \sqrt{n}\left[\theta_3^{*,[1]}(\hat{\theta}_1^{(n)}) - \theta_3^{*,[1]}(\theta_1^*)\right] = \frac{\partial \theta_3^*}{\partial \theta_1^*}\sqrt{n}\left[\hat{\theta}_1^{(n)} - \hat{\theta}_1^{(n)}\right] + o_P(1).$$

Now to show Equation (42) all we need to show is that

$$\sqrt{n}\left[\theta_3^*\big(\hat{\theta}_{1:2}^{(n)}\big) - \theta_3^{*,[1]}\big(\hat{\theta}_1^{(n)}\big)\right] = \frac{\partial \theta_3^*}{\partial \theta_2^*}\sqrt{n}\left[\hat{\theta}_2^{(n)}\big(\hat{\theta}_1^{(n)}\big) - \theta_2^*\big(\hat{\theta}_1^{(n)}\big)\right] + o_P(1).$$

By Condition 5, $\theta_3^{*,[1]}(\theta_1) = \theta_3^*(\theta_1, \theta_2^*(\theta_1))$ is Frechet differentiable with respect to $\theta_2^*(\theta_1)$ uniformly over $\theta_1 \in \Theta_1$, i.e., for any $\theta_2(\,\cdot\,) : \mathbb{R}^{d_1} \mapsto \mathbb{R}^{d_2}$,

$$\sup_{\theta_1 \in \Theta_1} \left\| \theta_3^*\big(\theta_1, \theta_2(\theta_1)\big) - \theta_3^*\big(\theta_1, \theta_2^*(\theta_1)\big) - \frac{\partial \theta_3^*\big(\theta_1, \tilde{\theta}_2\big)}{\partial \tilde{\theta}_2}\bigg|_{\tilde{\theta}_2 = \theta_2^*(\theta_1)} \big(\theta_2(\theta_1) - \theta_2^*(\theta_1)\big) \right\| = o\left(\sup_{\theta_1 \in \Theta_1} \big\|\theta_2(\theta_1) - \theta_2^*(\theta_1)\big\|\right).$$

Since $\mathbb{P}\big(\hat{\theta}_1^{(n)} \in \Theta_1\big) \to 1$ by consistency of $\hat{\theta}_1^{(n)}$, by the above Frechet differentiablity result,

$$\begin{aligned}\sqrt{n}\left[\theta_3^*\big(\hat{\theta}_{1:2}^{(n)}\big) - \theta_3^{*,[1]}\big(\hat{\theta}_1^{(n)}\big)\right] &= \sqrt{n}\left[\theta_3^*\big(\hat{\theta}_1^{(n)}, \hat{\theta}_2^{(n)}(\hat{\theta}_1^{(n)})\big) - \theta_3^*\big(\hat{\theta}_1^{(n)}, \theta_2^*(\hat{\theta}_1^{(n)})\big)\right] \\ &= \frac{\partial \theta_3^{*,[1]}(\hat{\theta}_1)}{\partial \theta_2^*(\hat{\theta}_1)}\sqrt{n}\left[\hat{\theta}_2^{(n)}\big(\hat{\theta}_1^{(n)}\big) - \theta_2^*\big(\hat{\theta}_1^{(n)}\big)\right] + \sqrt{n} o_P\left(\sup_{\theta_1 \in \Theta_1} \big\|\hat{\theta}_2^{(n)}(\theta_1) - \theta_2^*\big(\theta_1\big)\big\|\right) + o_P(1).\end{aligned}$$

By Lemma 8, we have that $\sqrt{n} o_P\left(\sup_{\theta_1 \in \Theta_1} \big\|\hat{\theta}_2^{(n)}\big(\theta_1\big) - \theta_2^*\big(\theta_1\big)\big\|\right) = o_P(1)$.

By Condition 5, $\frac{\partial \theta_3^{*,[1]}(\theta_1)}{\partial \theta_2^*(\theta_1)}$ is continuous in $\theta_1 \in \Theta_1$, so we can apply continuous mapping theorem to get $\frac{\partial \theta_3^{*,[1]}(\hat{\theta}_1)}{\partial \theta_2^*(\hat{\theta}_1)} \xrightarrow{P} \frac{\partial \theta_3^{*,[1]}(\theta_1^*)}{\partial \theta_2^*(\theta_1^*)} = \frac{\partial \theta_3^*}{\partial \theta_2^*}$.

By the above results, we have that $\sqrt{n}\big(\hat{\theta}_3^{(n)} - \theta_3^*\big) \xrightarrow{D} \mathcal{N}\left(0, \left[\frac{\partial \theta_t^*}{\partial \theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,s}(\dot{\Psi}_s^{-1})^\top \frac{\partial \theta_t^*}{\partial \theta_s^*}^\top\right]_{u=1,s=1}^{u=3,s=3}\right)$.

Thus, we have shown that Equation (36) holds for the $t = 3$ case, assuming the $t = 2$ case holds.

Showing Equation (36) Holds for General t (Induction Step): For our induction assumption we assume that for all $s \in [1\colon t-1]$,

$$\sqrt{n}\left[\hat{\theta}_s^{(n)} - \theta_s^*\right] \xrightarrow{D} \mathcal{N}\left(0, \left[\frac{\partial \theta_t^*}{\partial \theta_u^*}\dot{\Psi}_u^{-1}\Sigma_{u,q}(\dot{\Psi}_q^{-1})^\top \frac{\partial \theta_t^*}{\partial \theta_q^*}^\top\right]_{u=1,q=1}^{u=s,q=s}\right). \tag{43}$$

We now show that Equation (36) holds, i.e., that for any $s \in [1\colon t]$ that

$$\sqrt{n}\left[\theta_t^{*,[s]}\big(\hat{\theta}_{1:s}^{(n)}\big) - \theta_t^{*,[s-1]}\big(\hat{\theta}_{1:s-1}^{(n)}\big)\right] = \frac{\partial \theta_t^*}{\partial \theta_s^*}\sqrt{n}\left[\hat{\theta}_s^{(n)}(\hat{\theta}_{1:s-1}^{(n)}) - \theta_s^*(\hat{\theta}_{1:s-1}^{(n)})\right] + o_P(1).$$

By Condition 5, $\theta_t^{*,[s-1]}(\,\cdot\,)$ is Frechet differentiable with respect to $\theta_s^*(\,\cdot\,)$, i.e., for $\theta_s(\,\cdot\,) : \Theta_{1:s-1} \mapsto \Theta_s$,

$$\sup_{\theta_{1:s-1} \in \Theta_{1:s-1}} \Big\| \theta_t^{*,[s]}\big(\theta_{1:s-1}, \theta_s(\theta_{1:s-1})\big) - \theta_t^{*,[s-1]}\big(\theta_{1:s-1}\big) - \frac{\partial \theta_t^{*,[s-1]}\big(\theta_{1:s-1}\big)}{\partial \theta_s^*(\theta_{1:s-1})}\big(\theta_s\big(\theta_{1:s-1}\big) - \theta_s^*\big(\theta_{1:s-1}\big)\big) \Big\|$$

$$= o\left(\sup_{\theta_{1:s-1} \in \Theta_{1:s-1}} \left\| \theta_s(\theta_{1:s-1}) - \theta_s^*(\theta_{1:s-1}) \right\| \right).$$

Since $\mathbb{P}(\hat{\theta}_k^{(n)} \in \Theta_k) \to 1$ for all $k \in [1\colon s-1]$ by the results of Theorem 1, by the above Frechet differentiablity result,

$$\sqrt{n}\left[\theta_t^{*,[s]}(\hat{\theta}_{1:s}^{(n)}) - \theta_t^{*,[s-1]}(\hat{\theta}_{1:s-1}^{(n)})\right] = \sqrt{n}\left[\theta_t^{*,[s]}(\hat{\theta}_{1:s-1}^{(n)}, \hat{\theta}_s^{(n)}(\hat{\theta}_{1:s-1}^{(n)})) - \theta_t^{*,[s]}(\hat{\theta}_{1:s-1}^{(n)}, \theta_s^*(\hat{\theta}_{1:s-1}^{(n)}))\right]$$

$$= \frac{\partial \theta_t^{*,[s-1]}(\hat{\theta}_{1:s-1}^{(n)})}{\partial \theta_s^*(\hat{\theta}_{1:s-1}^{(n)})} \sqrt{n}\left[\hat{\theta}_s^{(n)}(\hat{\theta}_{1:s-1}^{(n)}) - \theta_s^*(\hat{\theta}_{1:s-1}^{(n)})\right]$$
$$+ \sqrt{n} o_P\left(\sup_{\theta_{1:s-1} \in \Theta_{1:s-1}} \left\| \hat{\theta}_s^{(n)}(\theta_{1:s-1}) - \theta_s^*(\theta_{1:s-1}) \right\| \right) + o_P(1).$$

By Lemma 8, we have that $\sqrt{n} o_P\left(\sup_{\theta_{1:s-1} \in \Theta_{1:s-1}} \left\| \hat{\theta}_s^{(n)}(\theta_{1:s-1}) - \theta_s^*(\theta_{1:s-1}) \right\|\right) = o_P(1)$. Also, by Condition 5, $\frac{\partial \theta_t^{*,[s-1]}(\theta_{1:s-1})}{\partial \theta_s^*(\theta_{1:s-1})}$ is continuous in $\theta_{1:s-1}$, so we can apply continuous mapping theorem to get $\frac{\partial \theta_t^{*,[s-1]}(\hat{\theta}_{1:s-1}^{(n)})}{\partial \theta_s^*(\hat{\theta}_{1:s-1}^{(n)})} \xrightarrow{P} \frac{\partial \theta_t^{*,[s-1]}(\theta_{1:s-1}^*)}{\partial \theta_s^*(\theta_{1:s-1}^*)} = \frac{\partial \theta_t^*}{\partial \theta_s^*}$.

Thus, we have shown that Equation (36) holds given our induction assumption Equation (43). By induction Equation (36) holds for general t. ■

C.4. Lemma 10: Importance-Weighted Martingale Central Limit Theorem

Lemma 10 (Importance-Weighted Martingale Central Limit Theorem) *We consider the problem setting as described in Section 1. Let $f_1, f_2, ..., f_T$ be real-valued functions of $\mathcal{H}_1^{(i)}, \mathcal{H}_2^{(i)}, ..., \mathcal{H}_T^{(i)'}$ respectively. We show that*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{f_1(\mathcal{H}_1^{(i)}) + \sum_{t=2}^{T} W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1}) f_t(\mathcal{H}_t^{(i)}) - \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t'=1}^{T} f_{t'}(\mathcal{H}_{t'}^{(i)})\right]\right\} \xrightarrow{D} \mathcal{N}\left(0, \mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right)\right)$$

for $\mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right) \triangleq \mathbb{E}_{\pi^*_{2:t}}\left[\left\{\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right\}^2\right] - \mathbb{E}_{\pi^*_{2:t}}\left[\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right]^2$ *assuming Conditions 1 and 2 and that*

(a) $\hat{\theta}_t^{(n)} - \theta_t^* = O_P(1/\sqrt{n})$ *for all* $t \in [1\colon T-1]$

(b) For some $\alpha > 0$, $\mathbb{E}_{\pi^*_{2:t}}\left[\left\|f\left(\mathcal{H}_t^{(i)}\right)\right\|_1^{4+\alpha}\right] < \infty$, *for all* $t \in [1\colon T]$.

Proof of Lemma 10 For convenience, we abbreviate $W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})$ as $W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})$. We first write the term of interest in the form of martingale differences:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) - \mathbb{E}\left[W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})\right]\right\}$$

Let $Y_t^{(i)} \triangleq W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})$ and $Y_{1:T}^{(i)} \triangleq \sum_{t=1}^{T} Y_t^{(i)}$.

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{Y_t^{(i)} - \mathbb{E}\left[Y_t^{(i)}\right]\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{Y_{1:T}^{(i)} - \mathbb{E}\left[Y_{1:T}^{(i)}\right]\right\}$$

We let $\mathcal{H}_0^{(1:n)} \triangleq \emptyset$ and $X_{T+1}^{(1:n)} \triangleq \emptyset$. Note that $Y_{1:T}^{(i)} = \mathbb{E}\left[Y_{1:T}^{(i)} | \mathcal{H}_T^{(1:n)}, X_{T+1}^{(1:n)}\right]$. By telescoping series,

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{\left\{\mathbb{E}\left[Y_{1:T}^{(i)} | \mathcal{H}_0^{(1:n)}, X_1^{(1:n)}\right] - \mathbb{E}\left[Y_{1:T}^{(i)}\right]\right\}}_{\triangleq Z_0^{(i)}} + \sum_{t=1}^{T}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{\left\{\mathbb{E}\left[Y_{1:T}^{(i)} | \mathcal{H}_t^{(1:n)}, X_{t+1}^{(1:n)}\right] - \mathbb{E}\left[Y_{1:T}^{(i)} | \mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right]\right\}}_{\triangleq Z_t^{(i)}}\right].$$

Note that $\mathbb{E}\left[Z_t^{(i)} | H_{t-1}^{(1:n)}, X_t^{(1:n)}\right] = 0$ for all $i \in [1\colon n]$ and $t \in [1\colon T]$.

In the next two subsections we will show the following two results:

(i) Convergence of conditional variance:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_0^{(i)}\right)^2\right]+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_t^{(i)}\right)^2\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]\xrightarrow{P}\mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T}f_t(\mathcal{H}_t^{(i)})\right). \quad (44)$$

(ii) Conditional Lindeberg: For any $\epsilon>0$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_0^{(i)}\right)^2\mathbb{I}_{|Z_0^{(i)}|/\sqrt{n}>\epsilon}\right]+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_t^{(i)}\right)^2\mathbb{I}_{|Z_t^{(i)}|/\sqrt{n}>\epsilon}\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]\xrightarrow{P}0. \quad (45)$$

With the above two results we can apply the central limit theorem for dependent random variables of Dvoretzky (1972) to conclude that our desired result holds, i.e.,

$$\frac{1}{\sqrt{n}}\sum_{t=0}^{T}\sum_{i=1}^{n}Z_t^{(i)}\xrightarrow{D}\mathcal{N}\left(0,\mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T}f_t(\mathcal{H}_t^{(i)})\right)\right).$$

C.4.1. CONDITIONAL VARIANCE

In this subsection, we show that Equation (44) holds. Using the definition of $Z_t^{(i)}$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_0^{(i)}\right)^2\right]+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(Z_t^{(i)}\right)^2\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_0^{(1:n)},X_1^{(1:n)}\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\right]\right)^2\right]$$

$$+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]\right)^2\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]. \quad (46)$$

Thus, we can simplify Equation (46) as follows:

$$=\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|X_1^{(1:n)}\right]^2\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\right]^2\right\}$$

$$+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]^2\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]^2\right\}.$$

Note by re-indexing, $\sum_{t=1}^{T}\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]^2=\sum_{t=0}^{T-1}\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]^2$
$=\sum_{t=1}^{T}\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]^2+\mathbb{E}\left[Y_{1:T}^{(i)}\middle|X_1^{(1:n)}\right]^2-\left(Y_{1:T}^{(i)}\right)^2$. By rearranging terms,

$$=\frac{1}{n}\sum_{i=1}^{n}\left\{\left(Y_{1:T}^{(i)}\right)^2-\mathbb{E}\left[Y_{1:T}^{(i)}\right]^2\right\}+\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|X_1^{(1:n)}\right]^2\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\middle|X_1^{(1:n)}\right]^2\right\}$$

$$+\sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]^2\middle|\mathcal{H}_{t-1}^{(1:n)},X_t^{(1:n)}\right]-\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)},X_{t+1}^{(1:n)}\right]^2\right\}. \quad (47)$$

Note the following observations

- We now show some helpful results (recall we abbreviate $W_{2:t}^{(i)}(\theta^*_{1:t-1}, \hat{\theta}^{(n)}_{1:t-1})$ as $W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})$):

$$\begin{aligned}\left|W_t^{(i)}(\theta^*, \hat{\theta}^{(n)}) - 1\right| &= \left|W_t^{(i)}(\theta^*, \hat{\theta}^{(n)}) - W_t^{(i)}(\theta^*, \theta^*)\right| \\ &\leq \left|\hat{\pi}_t\big(A_t^{(i)}, X_t^{(i)}\big)^{-1} - \pi_t^*\big(A_t^{(i)}, X_t^{(i)}\big)^{-1}\right| \leq \max_{a\in\mathcal{A}}\left|\hat{\pi}_t\big(a, X_t^{(i)}\big)^{-1} - \pi_t^*\big(a, X_t^{(i)}\big)^{-1}\right| \\ &\underbrace{\leq}_{(i)} \pi_{\min}^{-2} \max_{a\in\mathcal{A}}\left|\hat{\pi}_t\big(a, X_t^{(i)}\big) - \pi_t^*\big(a, X_t^{(i)}\big)\right| \underbrace{\leq}_{(ii)} \pi_{\min}^{-2} m_t\big(X_t^{(i)}\big)\big\|\hat{\theta}^{(n)}_{t-1} - \theta^*_{t-1}\big\|. \quad (48)\end{aligned}$$

Inequality (i) above holds because by Taylor Series expansion, $\hat{\pi}^{-1} - \pi^{*,-1} = (-1)\tilde{\pi}^{-2}(\hat{\pi} - \pi^*)$ for some $\tilde{\pi}$ between $\hat{\pi}$ and π^*, which means $\tilde{\pi} \geq \pi_{\min} > 0$ w.p. 1 by exploration Condition 1. Inequality (ii) holds by Lipschitz policy condition 2. Since $\big\|\hat{\theta}^{(n)}_{t-1} - \theta^*_{t-1}\big\| = O_P(1/\sqrt{n})$ by condition (a), we have that $W_t^{(i)}(\theta^*, \hat{\theta}^{(n)}) = 1 + O_P(1/\sqrt{n})$.

$$W_{2:t}^{(i)}\big(\theta^*, \hat{\theta}^{(n)}\big) = \big(1 + O_P(1/\sqrt{n})\big)^{t-1} = 1 + O_P(1/\sqrt{n}). \quad (49)$$

- We now show that for any real-valued function $g(\mathcal{H}_t^{(i)})$ with $\mathbb{E}_{\pi^*_{2:t}}\big[g(\mathcal{H}_t^{(i)})^2\big] < \infty$, then $g(\mathcal{H}_t^{(i)}) = O_P(1)$. Let $\epsilon > 0$. By Chebychev inequality and since $\pi_{\min}^{t-1} \leq \big(\frac{\pi_{\min}}{1-\pi_{\min}}\big)^{t-1} \leq W_{2:t}^{(i)}\big(\theta^*_{1:t-1}, \hat{\theta}_{1:t-1}\big)$, w.p. 1 by Condition 1,

$$\begin{aligned}\mathbb{P}\big(g(\mathcal{H}_t^{(i)}) > c_\epsilon\big) \leq c_\epsilon^{-2}\mathbb{E}\big[g(\mathcal{H}_t^{(i)})^2\big] &\leq c_\epsilon^{-2}\pi_{\min}^{-(t-1)}\mathbb{E}\big[W_{2:t}^{(i)}\big(\theta^*, \hat{\theta}^{(n)}\big)g(\mathcal{H}_t^{(i)})^2\big] \\ &= c_\epsilon^{-2}\pi_{\min}^{-(t-1)}\mathbb{E}_{\pi^*_{2:t}}\big[g(\mathcal{H}_t^{(i)})^2\big]. \quad (50)\end{aligned}$$

The above is less than ϵ by choosing $c_\epsilon > \sqrt{\epsilon^{-1}\pi_{\min}^{-(t-1)}\mathbb{E}_{\pi^*_{2:t}}\big[g(\mathcal{H}_t^{(i)})^2\big]}$.

- Since $Y_{1:T}^{(i)} = \sum_{t=1}^T Y_t^{(i)} = \sum_{t=1}^T W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})$,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Y_{1:T}^{(i)}\right]^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^T f_t(\mathcal{H}_t^{(i)})\right]^2 = \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^T f_t(\mathcal{H}_t^{(i)})\right]^2.$$

Also, note that

$$\frac{1}{n}\sum_{i=1}^n \big(Y_{1:T}^{(i)}\big)^2 = \frac{1}{n}\sum_{i=1}^n \left\{\sum_{t=1}^T W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})\right\}^2$$

$$= \frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) W_{2:s}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_s(\mathcal{H}_s^{(i)})$$

$$= \frac{1}{n}\sum_{i=1}^n \sum_{t=1}^T \sum_{s=1}^T W_{2:\max(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) f_s(\mathcal{H}_s^{(i)}) W_{2:\min(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)})$$

Note that since $\mathbb{E}_{\pi^*_{2:t}}[f_t(\mathcal{H}_t^{(i)})^2]$ is bounded by assumption, by Equation (50), $f_t(\mathcal{H}_t^{(i)}) = O_P(1)$. Thus, by Equation (49), we have that $W^{(i)}_{2:\min(t,s)}(\theta^*, \hat{\theta}^{(n)}) = 1 + O_P(1/\sqrt{n})$, so

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\sum_{s=1}^{T} W^{(i)}_{2:\max(t,s)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) f_s(\mathcal{H}_s^{(i)}) + O_P(1/\sqrt{n})\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{T} W^{(i)}_{2:\max(t,s)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) f_s(\mathcal{H}_s^{(i)}) + o_P(1).$$

By moment Condition (b), we can apply the Weighted Martingale Weak Law of Large Numbers (Lemma 7) to get

$$\xrightarrow{P} \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{T}\sum_{s=1}^{T} f_t(\mathcal{H}_t^{(i)}) f_s(\mathcal{H}_s^{(i)})\right] = \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right\}^2\right].$$

Since $\mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T} f(\mathcal{H}_t^{(i)})\right) = \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right\}^2\right] - \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\right]^2$, by the above results,

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y^{(i)}_{1:T}\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y^{(i)}_{1:T}\right]^2 \xrightarrow{P} \mathrm{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T} f(\mathcal{H}_t^{(i)})\right). \tag{51}$$

- We now consider the second summation term (out of three) in Equation (47). Note that $\mathbb{E}\left[\mathbb{E}\left[Y^{(i)}_{1:T}\middle|X_1^{(1:n)}\right]^2\right] = \mathbb{E}\left[\mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(i)}\right]^2\right]$. Also note the following:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y^{(i)}_{1:T}\middle|X_1^{(1:n)}\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\mathbb{E}\left[W^{(i)}_{2:t}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(1:n)}\right]\right\}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(1:n)}\right]\right\}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(i)}\right]\right\}^2$$

By the law of large numbers for i.i.d. random variables,

$$\xrightarrow{P} \mathbb{E}\left[\left\{\sum_{t=1}^{T}\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(i)}\right]\right\}^2\right] = \mathbb{E}\left[\mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|X_1^{(i)}\right]^2\right].$$

Thus, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y^{(i)}_{1:T}\middle|X_1^{(1:n)}\right]^2\right] - \mathbb{E}\left[Y^{(i)}_{1:T}\middle|X_1^{(1:n)}\right]^2\right\} \xrightarrow{P} 0. \tag{52}$$

By Equations (51) and (52) above, we have that Equation (47) equals the following:

$$= o_P(1) + \text{Var}_{\pi^*_{2:T}}\left(\sum_{t=1}^{T} f(\mathcal{H}_t^{(i)})\right)$$
$$+ \sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)}, X_{t+1}^{(1:n)}\right]^2\middle|\mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right] - \mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_t^{(1:n)}, X_{t+1}^{(1:n)}\right]^2\right\}.$$

The remainder of the proof in this section will be to show that for any $t' \in [1\colon T]$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right] - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]^2 \xrightarrow{P} 0. \quad (53)$$

We show that both summations above converge in probability to the same value.

Proof of Equation (53) Second Summation Note that for any $t' \in [1\colon T]$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\sum_{t=1}^{t'} Y_t^{(i)} + \sum_{t=t'+1}^{T} Y_t^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]^2$$

Since $Y_t^{(i)} \triangleq W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)})$,

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{t'} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) f_t(\mathcal{H}_t^{(i)}) + W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi^*_{t'+1:T}}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]\right)^2$$

For convenience, let $\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) \triangleq f_t(\mathcal{H}_t^{(i)})$ for all $t \in [1\colon t'-1]$ and let $\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t+1}^{(i)}) \triangleq f_{t'}(\mathcal{H}_{t'}^{(i)}) + \mathbb{E}_{\pi^*_{t'+1:T}}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]$.

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{t=1}^{t'} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right\}^2 \quad (54)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{t'}\sum_{s=1}^{t'} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) W_{2:s}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_s(\mathcal{H}_s^{(i)}, X_{s+1}^{(i)})$$

$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{t'}\sum_{s=1}^{t'} W_{2:\max(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\tilde{f}_s(\mathcal{H}_s^{(i)}, X_{s+1}^{(i)}) W_{2:\min(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)})$$

Note that $\mathbb{E}_{\pi^*_{2:t}}\left[f_t(\mathcal{H}_t^{(i)})^2\right]$ is bounded by assumption and since by Jensen's inequality $\mathbb{E}_{\pi^*_{2:t'}}\left[\mathbb{E}_{\pi^*_{t'+1:t}}\left[f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]^2\right] \leq \mathbb{E}_{\pi^*_{2:t}}[f_t(\mathcal{H}_t^{(i)})^2]$ is also bounded. Thus by Equation (50), we have that $\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) = O_P(1)$, where recall that for $t \leq t'$, $\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) = f_t(\mathcal{H}_t^{(i)})$ and $t = t'$, $\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}) = f_{t'}(\mathcal{H}_t^{(i)}) + \mathbb{E}_{\pi^*_{t'+1:T}}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]$.

Now, by Equation (49), $W_{2:\min(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)}) = 1 + O_P(1/\sqrt{n})$, so

$$= o_P(1) + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{t'}\sum_{s=1}^{t'} W_{2:\max(t,s)}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\tilde{f}_s(\mathcal{H}_s^{(i)}, X_{s+1}^{(i)})$$

By moment condition (b), we can apply the Weighted Martingale Weak Law of Large Numbers (Lemma 7) to get

$$\xrightarrow{P} \mathbb{E}_{\pi^*_{2:T}}\left[\sum_{t=1}^{t'}\sum_{s=1}^{t'}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\tilde{f}_s(\mathcal{H}_s^{(i)}, X_{s+1}^{(i)})\right] = \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\sum_{t=1}^{t'}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right\}^2\right] \tag{55}$$

$$= \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\sum_{t=1}^{t'} f_t(\mathcal{H}_t^{(i)}) + \mathbb{E}_{\pi^*_{t'+1:T}}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]\right\}^2\right]. \tag{56}$$

Proof of Equation (53) First Summation For any $t' \in [1\colon T]$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{E}\left[Y_{1:T}^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right].$$

By Equation (54) above,

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left\{\sum_{t=1}^{t'} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right\}^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left\{W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}) + \sum_{t=1}^{t'-1} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right\}^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]$$

$$= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})^2\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]}_{(i)}$$

$$+ \underbrace{2\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]\sum_{t=1}^{t'-1} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})}_{(ii)}$$

$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{t'-1} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right)^2}_{(iii)} \tag{57}$$

Term (iii): By the same argument made for Equation (54),

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{t'-1} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right)^2 \xrightarrow{P} \mathbb{E}_{\pi^*_{2:T}}\left[\left(\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right)^2\right] \tag{58}$$

Term (ii):

$$2\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]\sum_{t=1}^{t'-1} W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})$$

Since $W_{2:t}^{(i)}(\theta^*, \hat{\theta}^{(n)}) = 1 + O_P(1/\sqrt{n})$ by Equation (49) and since $\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) = O_P(1)$ (see text below Equation (54) for justification),

$$= 2\frac{1}{n}\sum_{i=1}^{n} W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(i)}, X_{t'}^{(i)}\right]\sum_{t=1}^{t'-1}\left\{\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) + O_P(1/\sqrt{n})\right\}$$

$$= o_P(1) + 2\frac{1}{n}\sum_{i=1}^{n} W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(i)}, X_{t'}^{(i)}\right]\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})$$

$$= o_P(1) + 2\frac{1}{n}\sum_{i=1}^{n} W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(i)}, X_{t'}^{(i)}\right].$$

By moment condition (b), we can apply the Weighted Martingale Weak Law of Large Numbers (Lemma 7) to get that

$$\xrightarrow{P} 2\mathbb{E}_{\pi_{2:T}^*}\left[\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\middle|\mathcal{H}_{t'-1}^{(i)}, X_{t'}^{(i)}\right]\right]$$

$$= 2\mathbb{E}_{\pi_{2:T}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right]. \quad (59)$$

Term (i):

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[W_{2:t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})^2\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)})^2\mathbb{E}_{\pi_{t'}^*}\left[W_{t'}^{(i)}(\theta^*, \hat{\theta})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right].$$

We now show that $\mathbb{E}_{\pi_{t'}^*}\left[W_{t'}^{(i)}(\theta^*, \hat{\theta})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right] = O_P(1)$. By Condition 1, $W_{t'}^{(i)}(\theta^*, \hat{\theta}) \leq \pi_{\min}^{-1}$ w.p. 1, so it is sufficient to show that $\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right] = O_P(1)$. Note that $\mathbb{E}_{\pi_{2:t}^*}\left[f_t(\mathcal{H}_t^{(i)})^4\right]$ is bounded by assumption and since by Jensen's inequality $\mathbb{E}_{\pi_{2:t'}^*}\left[\mathbb{E}_{\pi_{t'+1:t}^*}\left[f_t(\mathcal{H}_t^{(i)})^2\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]^2\right] \leq \mathbb{E}_{\pi_{2:t}^*}\left[f_t(\mathcal{H}_t^{(i)})^4\right]$ is also bounded. Thus by Equation (50), we can show that $\mathbb{E}_{\pi_{t'}^*}\left[\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right] = O_P(1)$, where recall that for $t \leq t'$, $\tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) = f_t(\mathcal{H}_t^{(i)})$ and $t = t'$, $\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}) = f_{t'}(\mathcal{H}_t^{(i)}) + \mathbb{E}_{\pi_{t'+1:T}^*}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)}\right]$.

Since $W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)}) = 1 + O_P(1/\sqrt{n})$ by Equation (49),

$$= o_P(1) + \frac{1}{n}\sum_{i=1}^{n} W_{2:t'-1}^{(i)}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi_{t'}^*}\left[W_{t'}^{(i)}(\theta^*, \hat{\theta}^{(n)})\tilde{f}_{t'}(\mathcal{H}_{t'}^{(i)}, X_{t'+1}^{(i)})^2\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right].$$

Since $W_{t'}^{(i)}(\theta^*, \hat{\theta}^{(n)}) = 1 + W_{t'}^{(i)}(\theta^*, \hat{\theta}^{(n)}) - 1$,

$$= o_P(1) + \frac{1}{n}\sum_{i=1}^{n} W^{(i)}_{2:t'-1}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(i)}_{t'-1}, X^{(i)}_{t'}\right]$$
$$+ \frac{1}{n}\sum_{i=1}^{n} W^{(i)}_{2:t'-1}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi^*_{t'}}\left[\left(W^{(i)}_{t'}(\theta^*, \hat{\theta}^{(n)}) - 1\right)\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(1:n)}_{t'-1}, X^{(1:n)}_{t'}\right]$$
$$= o_P(1) + \mathbb{E}_{\pi^*_{2:T}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2\right] \tag{60}$$

The above limit holds by the following observations:

- By moment condition (b), we can apply the Weighted Martingale Weak Law of Large Numbers (Lemma 7) to get that

$$\frac{1}{n}\sum_{i=1}^{n} W^{(i)}_{2:t'-1}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(i)}_{t'-1}, X^{(i)}_{t'}\right]$$
$$= o_P(1) + \mathbb{E}_{\pi^*_{2:T}}\left[\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(i)}_{t'-1}, X^{(i)}_{t'}\right]\right] = o_P(1) + \mathbb{E}_{\pi^*_{2:T}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2\right]$$

- We show that the following is $o_P(1)$:

$$\frac{1}{n}\sum_{i=1}^{n} W^{(i)}_{2:t'-1}(\theta^*, \hat{\theta}^{(n)})\mathbb{E}_{\pi^*_{t'}}\left[\left(W^{(i)}_{t'}(\theta^*, \hat{\theta}^{(n)}) - 1\right)\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(1:n)}_{t'-1}, X^{(1:n)}_{t'}\right].$$

  By exploration condition 1, $W^{(i)}_{2:t'-1}(\theta^*, \hat{\theta}^{(n)}) \leq \pi_{\min}^{t'-2}$ w.p. 1,

$$\leq \pi_{\min}^{-(t'-2)}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\pi^*_{t'}}\left[\left|W^{(i)}_{t'}(\theta^*, \hat{\theta}^{(n)}) - 1\right|\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(1:n)}_{t'-1}, X^{(1:n)}_{t'}\right].$$

  By Equation (48), $\left|W^{(i)}_{t}(\theta^*, \hat{\theta}^{(n)}) - 1\right| \leq \pi_{\min}^{-2} m_t(X^{(i)}_t)\left\|\hat{\theta}^{(n)}_{t-1} - \theta^*_{t-1}\right\|$ w.p. 1, so

$$\leq \pi_{\min}^{-(t'-2)}\frac{1}{n}\sum_{i=1}^{n}\pi_{\min}^{-2} m_{t'}(X^{(i)}_{t'})\left\|\hat{\theta}^{(n)}_{t'-1} - \theta^*_{t'-1}\right\|\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(1:n)}_{t'-1}, X^{(1:n)}_{t'}\right].$$

  Recall above we showed that $\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(1:n)}_{t'-1}, X^{(1:n)}_{t'}\right] = O_P(1)$ (see the beginning of this section on Term (i)). Since $\left\|\hat{\theta}^{(n)}_{t-1} - \theta^*_{t-1}\right\| = O_P(1/\sqrt{n})$ by condition (a),

$$= \pi_{\min}^{-(t'-2)}\frac{1}{n}\sum_{i=1}^{n}\pi_{\min}^{-2} m_t(X^{(i)}_t)O_P(1/\sqrt{n})\mathbb{E}_{\pi^*_{t'}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2 \middle| \mathcal{H}^{(i)}_{t'-1}, X^{(i)}_{t'}\right] = o_P(1).$$

Thus, by Equations (58), (59), (60) above we have that Equation (57) equals the following:

$$= o_P(1) + \mathbb{E}_{\pi^*_{2:T}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})^2\right] + 2\mathbb{E}_{\pi^*_{2:T}}\left[\tilde{f}_{t'}(\mathcal{H}^{(i)}_{t'}, X^{(i)}_{t'+1})\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1})\right]$$
$$+ \mathbb{E}_{\pi^*_{2:T}}\left[\left(\sum_{t=1}^{t'-1}\tilde{f}_t(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1})\right)^2\right]$$

$$= o_P(1) + \mathbb{E}_{\pi^*_{2:T}}\left[\left(\sum_{t=1}^{t'} \tilde{f}_t(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\right)^2\right]$$

$$= \mathbb{E}_{\pi^*_{2:T}}\left[\left\{\sum_{t=1}^{t'} f_t(\mathcal{H}_t^{(i)}) + \mathbb{E}_{\pi^*_{t'+1:T}}\left[\sum_{t=t'+1}^{T} f_t(\mathcal{H}_t^{(i)})\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]\right\}^2\right]. \tag{61}$$

Thus, Equation (53) holds by Equations (56) and (61) above.

C.4.2. CONDITIONAL LINDEBERG

For any $\epsilon > 0$, we show that the following conditional Lindeberg term is $o_P(1)$:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(Z_0^{(i)})^2\mathbb{I}_{|Z_0^{(i)}|/\sqrt{n}>\epsilon}\right] + \sum_{t=1}^{T}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(Z_t^{(i)})^2\mathbb{I}_{|Z_t^{(i)}|/\sqrt{n}>\epsilon}\middle|\mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right]$$

Note that for any $\alpha > 0$, $\mathbb{I}_{|Z|/\sqrt{n}>\epsilon} = \mathbb{I}_{|Z|/(\epsilon\sqrt{n})>1} = \mathbb{I}_{|Z|^\alpha/(\epsilon\sqrt{n})^\alpha>1} \le |Z|^\alpha/(\epsilon\sqrt{n})^\alpha$.
Thus we can upper bound the previous equation as follows:

$$\le \frac{1}{n(\epsilon\sqrt{n})^\alpha}\sum_{i=1}^{n}\mathbb{E}\left[\left|Z_0^{(i)}\right|^{2+\alpha}\right] + \sum_{t=1}^{T}\frac{1}{n(\epsilon\sqrt{n})^\alpha}\sum_{i=1}^{n}\mathbb{E}\left[\left|Z_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right]. \tag{62}$$

Note that for any $\eta > 0$ and any numbers a, b, we have that $|a-b|^\eta \le c_\eta|a|^\eta + c_\eta|b|^\eta$ for some constant $c_\eta < \infty$. This c_η exists because

$$|a-b|^\eta \le (|a|+|b|)^\eta \le \begin{cases} (|a|+|b|)^{\lfloor\eta\rfloor} & \text{if } |a|+|b| \le 1 \\ (|a|+|b|)^{\lceil\eta\rceil} & \text{if } |a|+|b| > 1 \end{cases}$$

and for any positive integer k, by the Binomial theorem $(|a|+|b|)^k = \sum_{j=0}^{k}\binom{k}{j}\left(|a|^j + |b|^{k-j}\right) \le \left\{\sum_{j=0}^{k}\binom{k}{j}\right\}\left(|a|^k + |b|^k\right)$. Thus we can choose $c_\eta = \left\{\sum_{j=0}^{k}\binom{k}{j}\right\}$.

The above implies the following inequality for any numbers $a_1, a_2, ..., a_K$, $|\sum_{k=1}^{K} a_k|^\eta \le c_\eta^K \sum_{k=1}^{K}|a_k|^\eta$.
Thus, we have that for $t' \in [1\colon T]$,

$$\left|Z_{t'}^{(i)}\right|^{2+\alpha} = \left|\sum_{t=1}^{T}\left\{\mathbb{E}\left[Y_t^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right] - \mathbb{E}\left[Y_t^{(i)}\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]\right\}\right|^{2+\alpha}$$

$$\le c_{2+\alpha}^{2T}\sum_{t=1}^{T}\left\{\left|\mathbb{E}\left[Y_t^{(i)}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right]\right|^{2+\alpha} + \left|\mathbb{E}\left[Y_t^{(i)}\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]\right|^{2+\alpha}\right\}.$$

By Jensen's Inequality,

$$\le c_{2+\alpha}^{2T}\sum_{t=1}^{T}\left\{\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'}^{(1:n)}, X_{t'+1}^{(1:n)}\right] + \mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)}, X_{t'}^{(1:n)}\right]\right\}.$$

Thus, we can upper bound Equation (62) as follows:

$$\frac{c_{2+\alpha}^{2T}}{n(\epsilon\sqrt{n})^{\alpha}}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_0^{(1:n)},X_1^{(1:n)}\right]+\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\right]\right]$$

$$+\sum_{t'=1}^{T}\frac{c_{2+\alpha}^{2T}}{n(\epsilon\sqrt{n})^{\alpha}}\sum_{i=1}^{n}\sum_{t=1}^{T}\mathbb{E}\left[\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'}^{(1:n)},X_{t'+1}^{(1:n)}\right]+\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]$$

$$=\frac{c_{2+\alpha}^{2T}}{n(\epsilon\sqrt{n})^{\alpha}}\sum_{i=1}^{n}\sum_{t=1}^{T}2\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\right]+\sum_{t'=1}^{T}\frac{c_{2+\alpha}^{2T}}{n(\epsilon\sqrt{n})^{\alpha}}\sum_{i=1}^{n}\sum_{t=1}^{T}2\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]$$

To show that the above is $o_P(1)$, it is sufficient to show that the terms $\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]$, $\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\right]$ for all $t,t'\in[1\colon T]$ are all $O_P(1)$. By Chebychev inequality, to show that some random variable Y is $O_P(1)$ it is sufficient to show that $\mathbb{E}\big[Y^2\big]<\infty$. By Jensen's inequality,

$$\mathbb{E}\left[\mathbb{E}\left[\left|Y_t^{(i)}\right|^{2+\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]^2\right]\leq\mathbb{E}\left[\mathbb{E}\left[\left|Y_t^{(i)}\right|^{4+2\alpha}\middle|\mathcal{H}_{t'-1}^{(1:n)},X_{t'}^{(1:n)}\right]\right]\leq\mathbb{E}\left[\left|Y_t^{(i)}\right|^{4+2\alpha}\right].$$

Thus, it is sufficient to show that $\mathbb{E}\left[\left|Y_t^{(i)}\right|^{4+2\alpha}\right]<\infty$ for all $t\in[1\colon T]$. Since $W_{2:t}^{(i)}(\theta^*,\hat{\theta}^{(n)})\leq\pi_{\min}^{-(t-1)}$ w.p. 1 by exploration condition 1,

$$\mathbb{E}\left[\left|Y_t^{(i)}\right|^{4+2\alpha}\right]=\mathbb{E}\left[\left|W_{2:t}^{(i)}(\theta^*,\hat{\theta}^{(n)})f_t(\mathcal{H}_t^{(i)})\right|^{4+2\alpha}\right]$$

$$\leq\pi_{\min}^{-T(3+2\alpha)}\mathbb{E}\left[W_{2:t}^{(i)}(\theta^*,\hat{\theta}^{(n)})\left|f_t(\mathcal{H}_t^{(i)})\right|^{4+2\alpha}\right]=\pi_{\min}^{-T(3+2\alpha)}\mathbb{E}_{\pi_{2:t}^*}\left[\left|f_t(\mathcal{H}_t^{(i)})\right|^{4+2\alpha}\right]<\infty.$$

The above is bounded by moment condition (b) and minimum exploration condition 1. ■

Appendix D. Asymptotic Equicontinuity

We now provide an overview of the main results we prove in this section. First though, we introduce some notation that we use throughout this section. For notational convenience, we let $\theta_T \triangleq \theta$ and $\theta_t \triangleq \beta_t$ for all $t \in [1\colon T-1]$. This means that $\theta_T^* \triangleq \theta^*$, $\theta_t^* \triangleq \beta_t^*$, $\hat{\theta}_T^{(n)} \triangleq \hat{\theta}^{(n)}$, and $\hat{\theta}_t^{(n)} \triangleq \hat{\beta}_t^{(n)}$. Also we use $\Theta_T \triangleq \Theta$ and $\Theta_t \triangleq B_t$, where recall Θ is a bounded ball that contains $\theta^*(\beta_{1:T-1})$ for all $\beta_{1:T-1} \in B_{1:T-1}$ and B_t is a bounded ball that contains $\beta_t^*(\beta_{1:t-1})$ for all $\beta_{1:t-1} \in B_{1:t-1}$. Additionally, we let $\psi_T \triangleq \psi$ and $\psi_t \triangleq \phi_t$ for all $t \in [1\colon T-1]$, so $\psi_T(\mathcal{H}_T^{(i)}; \theta_T) = \psi(\mathcal{H}_T^{(i)}; \theta)$ and $\psi_t(\mathcal{H}_t^{(i)}; \theta_t) = \phi_t(\mathcal{H}_t^{(i)}; \beta_t)$ for $t < T$. Throughout this appendix $\mathbb{E}$, without a subscript, indicates expectation with respect to the data generating distribution ($\hat{\pi}_{2:T}$ used to select actions).

Using our newly defined notation, for any $t \in [1\colon T]$, we define the following functions of $\theta_1 \in \mathbb{R}^{d_1}$, $\theta_2 \in \mathbb{R}^{d_2}, \ldots, \theta_t \in \mathbb{R}^{d_t}$:

$$\Psi_t(\theta_{1:t}) \triangleq \mathbb{E}\left[\left\{\prod_{s=2}^{t} W_s^{(i)}\big(\theta_{s-1}, \hat{\theta}_{s-1}^{(n)}\big)\right\}\psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big)\right]$$

Above, $\theta_{1:t} \triangleq [\theta_1, \theta_2, \ldots, \theta_t]$. Recall that $W_s^{(i)}\big(\theta_{s-1}, \hat{\theta}_{s-1}^{(n)}\big) \triangleq \frac{\pi_s(A_s^{(i)}, X_s^{(i)}; \theta_{s-1})}{\pi_s(A_s^{(i)}, X_s^{(i)}; \hat{\theta}_{s-1}^{(n)})}$. Since the weights we use are importance weights, this means that

$$\Psi_t(\theta_{1:t}) = \mathbb{E}\left[\left\{\prod_{s=2}^{t} W_s^{(i)}\big(\theta_{s-1}, \hat{\theta}_{s-1}^{(n)}\big)\right\}\psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big)\right] = \mathbb{E}_{\pi_2(\theta_1), \pi_3(\theta_2), \ldots, \pi_t(\theta_{t-1})}\left[\psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big)\right].$$

Above, the expectation on the right is with respect to the distribution in which policies $\pi_2(\theta_1), \pi_3(\theta_2), \ldots, \pi_t(\theta_{t-1})$ are used to select actions. Further note that the above equality implies that $\Psi_t(\theta_{1:t})$ does not depend on the sample size, n.

We also define empirical versions of the above functions $\Psi_t(\theta_{1:t})$. Specifically, for any $t \in [1\colon T]$, we define:

$$\hat{\Psi}_t^{(n)}(\theta_{1:t}) \triangleq \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{s=2}^{t} W_s^{(i)}\big(\theta_{s-1}, \hat{\theta}_{s-1}^{(n)}\big)\right\}\psi_t\big(\mathcal{H}_t^{(i)}; \theta_t\big).$$

Recall that as discussed in the main text, θ_t^* and $\hat{\theta}_t^{(n)}$ are Z-estimators and can be considered implicitly defined functions of $\theta_1 \in \mathbb{R}^{d_1}$, $\theta_2 \in \mathbb{R}^{d_2}$, $\ldots$, $\theta_{t-1} \in \mathbb{R}^{d_{t-1}}$. Specifically, $\theta_t^*(\theta_{1:t-1})$ and $\hat{\theta}_t^{(n)}(\theta_{1:t-1})$ respectively satisfy

$$0 = \Psi_t\big(\theta_{1:t}\big)\big|_{\theta_t = \theta_t^*(\theta_{1:t-1})} \qquad \text{and} \qquad 0 = \hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big)\big|_{\theta_t = \hat{\theta}_t^{(n)}(\theta_{1:t-1})}.$$

For each $t \in [1\colon T]$, we overload notation by using $\theta_t^*(\,\cdot\,), \hat{\theta}_t^{(n)}(\,\cdot\,)$ to refer to the functions of $\theta_{1:t-1}$, and use $\theta_t^*, \hat{\theta}_t^{(n)}$ to refer to the vectors $\theta_t^*(\theta_{1:t-1}^*), \hat{\theta}_t^{(n)}\big(\hat{\theta}_{1:t-1}^{(n)}\big)$.

With the above defined notation in hand, we now describe the main result we prove in this section, which is that for any $t \in [1\colon T]$, for any fixed $c_t \in \mathbb{R}^{d_t}$,

$$\Bigg\| \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,)\big) \right] - \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\,\cdot\,, \theta_t^*(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \theta_t^*(\,\cdot\,)\big) \right] \Bigg\|_{\Theta_{1:t-1}} \xrightarrow{P} 0. \quad (63)$$

Recall that $\|f(\,\cdot\,)\|_{\Theta_{1:t-1}} = \sup_{\theta_{1:t-1} \in \Theta_{1:t-1}} \|f(\,\cdot\,)\|$.

Specifically, Lemma 16 below will prove Equation (63). The two key results that Lemma 16 uses are that $\|\hat{\theta}_t^{(n)}(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{\Theta_{1:t}} \xrightarrow{P} 0$ for all $t \in [1\colon T]$ (Theorem 1) and that the following stochastic process is asymptotically tight (specifically we will show that it is functionally asymptotically Gaussian):

$$\left\{ \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big) - \Psi_t\big(\theta_{1:t}\big) \right] : \theta_s \in \Theta_s \text{ for all } s \in [1\colon t] \right\}. \quad (64)$$

Note that for the following class of functions (first introduced below Condition 4, but here using θ_t notation instead of β_t notation) for each $t \in [1\colon T]$ and any $c_t \in \mathbb{R}^{d_t}$:

$$\mathcal{F}_{t,c_t} \triangleq \left\{ \left(\prod_{s=2}^{t-1} \pi_s(\,\cdot\,;\theta_{s-1}) \right) c_t^\top \psi_t\big(\,\cdot\,;\theta_t\big) : \theta_s \in \Theta_s \text{ for all } s \in [1\colon t] \right\}. \quad (65)$$

Note that by our careful choice of $\mathcal{F}_{t,c_t}$ above, that the stochastic process in Equation (64) is *equivalent* to the following stochastic process:

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1} f(\mathcal{H}_t^{(i)}) - \mathbb{E}\big[\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1} f(\mathcal{H}_t^{(i)}) \big] \right) : f \in \mathcal{F}_{t,c_t} \right\}. \quad (66)$$

Above, we use $\hat{\pi}_t^{(i)} \triangleq \hat{\pi}_t\big(A_t^{(i)}, X_t^{(i)}\big)$ and $\hat{\pi}_{2:t}^{(i)} \triangleq \prod_{s=2}^{t} \hat{\pi}_s^{(i)}$. Similarly, we will also use $\pi_s^{*,(i)} \triangleq \pi_t^*\big(A_t^{(i)}, X_t^{(i)}\big)$ and $\pi_{2:t}^{*,(i)} \triangleq \prod_{s=2}^{t} \pi_s^{*,(i)}$.

Note that since for any time t, $\{\mathcal{H}_t^{(i)}\}_{i=1}^n$ are not independent in our setting, classical empirical process theory for i.i.d. data cannot be used to prove that the stochastic process in Equation (66) is asymptotically tight in $l^\infty(\mathcal{F}_{t,c_t})$ (the collection of all bounded functions from $\mathcal{F}_{t,c_t}$ to $\mathbb{R}$). We now provide a summary of results in this section and describe the ways in which our results are similar to and differ from classical results in empirical processes.

Summary of Results in this Section

- **Lemma 11** proves a Bernstein inequality for our non-independent data type and is the most novel step in this section. The proof leverages the conditional independence of the action selection at each time-step and the fact that the underlying potential outcomes are i.i.d. The proof repeatedly uses a key helper Lemma 12.
- **Lemma 13** proves a maximal inequality for stochastic processes in the form of Equation (66) in the case that $|\mathcal{F}_{t,c_t}| < \infty$. The proof closely follows that of Lemma 19.33 (Van der Vaart, 2000), but replaces the use of a Bernstein inequality for i.i.d. data with Lemma 11.

- **Lemma 14** proves a maximal inequality for the stochastic process in Equation (66) as a function of the bracketing integral of class $\mathcal{F}_{t,c_t}$. The proof closely follows that of Lemma 19.34 (Van der Vaart, 2000), but replaces the use of a maximal inequality for empirical processes for a finite class of functions on i.i.d. data with Lemma 13.
- **Theorem 15** proves the stochastic process in Equation (66) is functionally asymptotically normal under a finite bracketing integral condition on $\mathcal{F}_{t,c_t}$. The proof closely follows that of Lemma 19.34 (Van der Vaart, 2000), but replaces a maximal inequality for empirical processes for function classes with finite bracketing integrals on i.i.d. data with Lemma 14.
- **Lemma 16** proves Equation (63) using Theorem 15 with an argument similar to Lemma 19.24 (Van der Vaart, 2000).

D.1. Lemma 11: Weighted Martingale Bernstein Inequality

Lemma 11 (Weighted Martingale Bernstein Inequality) *We consider the problem setting as described in Section 1. We assume Condition 1 holds. Let f be a real-valued function of $\mathcal{H}_T^{(i)}$ with $0 < \|f\|_\infty < \infty$. Then, for any $x > 0$ and for any $n \geq 1$,*

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) \right] \right\} > x \right) \leq \exp\left(-\frac{\pi_{\min}^{T-1}}{4} \frac{x^2}{\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})^2 \right] + x\|f\|_\infty/\sqrt{n}} \right).$$

Remarks Note that regarding the expectation on the right hand side in Lemma 11, for any fixed policies $\pi_{2:T}(\theta_{1:T-1})$,

$$\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})^2 \right] = \mathbb{E}_{\pi_{2:T}(\theta_{1:T-1})}\left[\left(\prod_{t=2}^{T} \pi_t(A_t^{(i)}, X_{t,i}; \theta_t) \right)^{-1} f(\mathcal{H}_T^{(i)})^2 \right].$$

Proof of Lemma 11 (Weighted Martingale Bernstein Inequality) We follow an argument similar to Lemma 19.32 in Van der Vaart (2000).

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) \right] \right\} > x \right) \tag{67}$$

By Chernoff bound for any $\lambda > 0$,

$$\leq e^{-\lambda x}\mathbb{E}\left[\exp\left\{ \frac{\lambda}{\sqrt{n}} \sum_{i=1}^{n} \left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) \right] \right) \right\} \right]$$

Changing the summation in exponent into a product,

$$= e^{-\lambda x}\mathbb{E}\left[\prod_{i=1}^{n} \exp\left\{ \frac{\lambda}{\sqrt{n}} \left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) \right] \right) \right\} \right]$$

We now apply Maclaurin series for exponential function, i.e., that $e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$.

$$= e^{-\lambda x}\mathbb{E}\left[\prod_{i=1}^{n}\sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{\lambda}{\sqrt{n}}\right)^k\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)^k\right]$$

Simplifying the first two terms in the inner summation,

$$= e^{-\lambda x}\mathbb{E}\Bigg[\prod_{i=1}^{n}\Bigg\{1 + \frac{\lambda}{\sqrt{n}}\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right) + \sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{\lambda}{\sqrt{n}}\right)^k\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)^k\Bigg\}\Bigg] \quad (68)$$

Note the following observations:

- In Equation (68), each of the terms in the product over n terms is non-negative because above we derived the product from $\prod_{i=1}^{n}\exp\left\{\frac{\lambda}{\sqrt{n}}\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)\right\}$ and $e^x \geq 0$ for all x.
- Since $\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} \leq \pi_{\min}^{-(T-1)}$ w.p. 1 by condition 1, $\left|\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right|$ $\leq \left|\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right| + \left|\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right| \leq 2\pi_{\min}^{-(T-1)}\|f\|_\infty$.
- We can upper bound the following:

$$\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)^2$$

$$= \left(\hat{\pi}_{2:T}^{(i)}\right)^{-2}f(\mathcal{H}_T^{(i)})^2 - 2\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right] + \left(\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)^2$$

Since $\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} \leq \pi_{\min}^{-(T-1)}$ w.p. 1 by condition 1,

$$\leq \left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}\pi_{\min}^{-(T-1)}f(\mathcal{H}_T^{(i)})^2 - 2\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right] + \left(\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right)^2$$

Note that $\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right] = \mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]$. For $g(\mathcal{H}_T^{(i)}) \triangleq \pi_{\min}^{-(T-1)}f(\mathcal{H}_T^{(i)})^2 - 2f(\mathcal{H}_T^{(i)})\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]$,

$$= \left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}g(\mathcal{H}_T^{(i)}) + \mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]^2$$

By the above observations, we can upper bound Equation (68) as follows:

$$e^{-\lambda x}\mathbb{E}\Bigg[\prod_{i=1}^{n}\Bigg\{1 + \frac{\lambda}{\sqrt{n}}\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right) + \sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{\lambda}{\sqrt{n}}\right)^k\left(\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}g(\mathcal{H}_T^{(i)}) + \mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]^2\right)\left(2\pi_{\min}^{-(T-1)}\|f\|_\infty\right)^{k-2}\Bigg\}\Bigg] \quad (69)$$

Note that everything in the expectation above in Equation (69) is bounded w.p. 1; we will show that this is true for the infinite summation over k. Let $y = \left(\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} g(\mathcal{H}^{(i)}_T) + \mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \right)$ and $z = 2\pi_{\min}^{-(T-1)} \|f\|_\infty$. Note that both y and z are bounded w.p. 1. Thus, since $\|f\|_\infty > 0$ by assumption, we have that $z > 0$, so $\sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k y z^{k-2} = y z^{-2} \sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k z^k \leq y z^{-2} e^{z\lambda/\sqrt{n}}$ is also bounded w.p. 1.

Moreover, Equation (69) can be written as $e^{-\lambda x} \mathbb{E}\left[\prod_{i=1}^n \left\{ \left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} h(\mathcal{H}^{(i)}_T) + c \right\} \right]$, for some function h and some finite constant c (i.e., c is a constant with respect to the index i and is non-random). Thus, we can apply Lemma 12 to get that Equation (69) is equal to

$$
\begin{aligned}
&= e^{-\lambda x} \prod_{i=1}^n \mathbb{E}_{\pi^*_{2:T}} \Bigg[1 + \frac{\lambda}{\sqrt{n}} \left(\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) - \mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right] \right) \\
&+ \sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k \left(\left(\pi^{*,(i)}_{2:T}\right)^{-1} g(\mathcal{H}^{(i)}_T) + \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \right) \left(2\pi_{\min}^{-(T-1)} \|f\|_\infty \right)^{k-2} \Bigg]
\end{aligned}
$$

Since $\mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right] = \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]$, we can cancel terms in the first line above.

$$
\begin{aligned}
&= e^{-\lambda x} \prod_{i=1}^n \Bigg\{ 1 + \mathbb{E}_{\pi^*_{2:T}} \Bigg[\sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k \left(\left(\pi^{*,(i)}_{2:T}\right)^{-1} g(\mathcal{H}^{(i)}_T) + \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \right) \\
&\qquad\qquad \left(2\pi_{\min}^{-(T-1)} \|f\|_\infty \right)^{k-2} \Bigg] \Bigg\}
\end{aligned}
$$

Since everything in the expectations above are bounded w.p. 1 (discussed below Equation (69)), we can exchange the expectation with the infinite summation over k.

$$
\begin{aligned}
&= e^{-\lambda x} \prod_{i=1}^n \Bigg\{ 1 + \sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k \left(\mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} g(\mathcal{H}^{(i)}_T) \right] + \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \right) \\
&\qquad\qquad \left(2\pi_{\min}^{-(T-1)} \|f\|_\infty \right)^{k-2} \Bigg\} \qquad (70)
\end{aligned}
$$

Since $g(\mathcal{H}^{(i)}_T) \triangleq \pi_{\min}^{-(T-1)} f(\mathcal{H}^{(i)}_T)^2 - 2 f(\mathcal{H}^{(i)}_T) \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]$,

$$
\begin{aligned}
&\mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} g(\mathcal{H}^{(i)}_T) \right] + \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \\
&= \pi_{\min}^{-(T-1)} \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)^2 \right] - 2\mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 + \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \\
&= \pi_{\min}^{-(T-1)} \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)^2 \right] - \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) \right]^2 \leq \pi_{\min}^{-(T-1)} \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)^2 \right]
\end{aligned}
$$

Thus Equation (70) can be upper bounded by the following:

$$
e^{-\lambda x} \prod_{i=1}^n \left\{ 1 + \sum_{k=2}^\infty \frac{1}{k!} \left(\frac{\lambda}{\sqrt{n}}\right)^k \pi_{\min}^{-(T-1)} \mathbb{E}_{\pi^*_{2:T}} \left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)^2 \right] \left(2\pi_{\min}^{-(T-1)} \|f\|_\infty \right)^{k-2} \right\}
$$

By i.i.d. potential outcomes,

$$= e^{-\lambda x}\left\{1+\sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{\lambda}{\sqrt{n}}\right)^k \pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]\left(2\pi_{\min}^{-(T-1)}\|f\|_\infty\right)^{k-2}\right\}^n$$

By rearranging terms,

$$= e^{-\lambda x}\left\{1+\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\lambda^k \underbrace{2\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}_{\triangleq\lambda_1^{-1}}\left(\underbrace{2\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}}_{\triangleq\lambda_2^{-1}}\right)^{k-2}\right\}^n$$

$$= e^{-\lambda x}\left\{1+\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\lambda^k\left(\lambda_1^{-1}\lambda_2^{-(k-2)}\right)\right\}^n \tag{71}$$

Note that since $\lambda_1^{-1}, \lambda_2^{-1} > 0$,

$$\lambda \triangleq x\left(\lambda_1^{-1}+x\lambda_2^{-1}\right)^{-1} \le \min\left(x\left(\lambda_1^{-1}+0\right)^{-1}, x\left(0+x\lambda_2^{-1}\right)^{-1}\right) = \min\left(x\lambda_1, \lambda_2\right).$$

Thus we have that $\lambda^k \le \lambda\min(x\lambda_1,\lambda_2)^{k-1} \le \lambda x\lambda_1\lambda_2^{k-2}$. So we can upper bound Equation (71) as follows:

$$\le e^{-\lambda x}\left\{1+\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\left(\lambda x\lambda_1\lambda_2^{k-2}\right)\left(\lambda_1^{-1}\lambda_2^{-(k-2)}\right)\right\}^n$$

$$= e^{-\lambda x}\left\{1+\frac{1}{n}\underbrace{\sum_{k=2}^{\infty}\frac{1}{k!}}_{\le 1}\frac{1}{2}\lambda x\right\}^n$$

By the Maclaurin series for exponential function, $e^z = \sum_{k=0}^{\infty}\frac{z^k}{k!}$, we have $\sum_{k=2}^{\infty}\frac{1}{k!} = e - \frac{1}{0!} - \frac{1}{1!} = e-2 \le 1$.

$$\le e^{-\lambda x}\left\{1+\frac{1}{n}\frac{1}{2}x\lambda\right\}^n$$

Again by the Maclaurin series for exponential function, $e^z = \sum_{k=0}^{\infty}\frac{z^k}{k!}$, so for $z > 0$ we have that $1+z \le e^z$, which means that $(1+z)^n \le e^{zn}$.

$$\le e^{-\lambda x}\exp\left(\frac{1}{2}x\lambda\right) = \exp\left(-\frac{1}{2}x\lambda\right)$$

Recall that $\lambda \triangleq x\left(\lambda_1^{-1}+x\lambda_2^{-1}\right)^{-1}$, so,

$$= \exp\left(-\frac{1}{2}xx\left(\lambda_1^{-1}+x\lambda_2^{-1}\right)^{-1}\right)$$

Recall that $\lambda_1^{-1} = 2\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]$ and $\lambda_2^{-1} = 2\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}$.

$$= \exp\left(-\frac{\pi_{\min}^{T-1}}{4}\frac{x^2}{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right] + x\|f\|_\infty/\sqrt{n}}\right). \blacksquare$$

Lemma 12 (Conditional Independence using Weights) *Let f be any real valued function of $\mathcal{H}_T^{(i)}$ such that $\mathbb{E}_{\pi^*_{2:T}}\left[f(\mathcal{H}_T^{(i)})\right] < \infty$. Let c be a non-random constant. Then the following equality holds:*

$$\mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)}) + c\right\}\right] = \prod_{i=1}^{n}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\prod_{t=2}^{T}\pi^*_t(A_t^{(i)}, X_t^{(i)})\right)^{-1} f(\mathcal{H}_T^{(i)}) + c\right].$$

Remarks Note that regarding the expectation terms on the right hand side above, for any fixed policies $\pi_{2:T}(\theta_{1:T-1})$,

$$\mathbb{E}_{\pi^*_{2:T}}\left[\left(\prod_{t=2}^{T}\pi^*_t(A_t^{(i)}, X_t^{(i)})\right)^{-1} f(\mathcal{H}_T^{(i)})\right] = \mathbb{E}_{\pi_{2:T}(\theta_{1:T-1})}\left[\left(\prod_{t=2}^{T}\pi_t(A_t^{(i)}, X_t^{(i)}; \theta_{t-1})^{-1}\right) f(\mathcal{H}_T^{(i)})\right]$$

Proof of Lemma 12 (Conditional Independence using Weights) We can show that for any $t \in [2:T]$, for any function g of $\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}$ that

$$\mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:t}\right)^{-1} g(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}) + c\right\}\right]$$
$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:t-1}\right)^{-1}\mathbb{E}_{\pi^*_t}\left[\left(\pi_t^{*,(i)}\right)^{-1} g(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)})\middle|\mathcal{H}_{t-1}^{(i)}, X_t^{(i)}\right] + c\right\}\right], \quad (72)$$

where $\pi_t^{*,(i)} \triangleq \pi^*_t(A_t^{(i)}, X_t^{(i)})$.

For now we take Equation (72) as given (see below for the proof). We will show that the desired result holds by repeatedly applying Equation (72). Applying Equation (72) for $t = T$ and $g = f$, we have that

$$\mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)}) + c\right\}\right] = \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T-1}\right)^{-1}\mathbb{E}_{\pi^*_T}\left[\left(\pi_T^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\middle|\mathcal{H}_{T-1}^{(i)}, X_T^{(i)}\right] + c\right\}\right].$$

Note that $\mathbb{E}_{\pi^*_T}\left[\left(\pi_T^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\middle|\mathcal{H}_{T-1}^{(i)}, X_T^{(i)}\right]$ is a function of $\mathcal{H}_{T-1}^{(i)}, X_T^{(i)}$; let this be function be g when we apply Equation (72) again for $t = T-1$.

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T-2}\right)^{-1}\mathbb{E}_{\pi^*_{T-1}}\left[\left(\pi_{T-1}^{*,(i)}\right)^{-1}\mathbb{E}_{\pi^*_T}\left[\left(\pi_T^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\middle|\mathcal{H}_{T-1}^{(i)}, X_T^{(i)}\right]\middle|\mathcal{H}_{T-2}^{(i)}, X_{T-1}^{(i)}\right] + c\right\}\right]$$

By law of iterated expectations,

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T-2}\right)^{-1}\mathbb{E}_{\pi^*_{T-1:T}}\left[\left(\pi_{T-1:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\middle|\mathcal{H}_{T-2}^{(i)}, X_{T-1}^{(i)}\right] + c\right\}\right].$$

By repeatedly applying Equation (72) for $t = T-2, T-3, ..., 2$ we have that

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T-3}\right)^{-1}\mathbb{E}_{\pi^*_{T-2:T}}\left[\left(\pi_{T-2:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\middle|\mathcal{H}_{T-3}^{(i)}, X_{T-2}^{(i)}\right] + c\right\}\right]$$

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:T-4}\right)^{-1}\mathbb{E}_{\pi^*_{T-3:T}}\left[\left(\pi^{*,(i)}_{T-3:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)\middle|\mathcal{H}^{(i)}_{T-4}, X^{(i)}_{T-3}\right] + c\right\}\right]$$

$$= \ldots = \mathbb{E}\left[\prod_{i=1}^{n}\left\{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)\middle|\mathcal{H}^{(i)}_1, X^{(i)}_1\right] + c\right\}\right].$$

Finally, recall that $\{\mathcal{H}^{(i)}_1, X^{(i)}_2\}^n_{i=1} = \{X^{(i)}_1, A^{(i)}_1, R^{(i)}_1, X^{(i)}_2\}^n_{i=1}$ are independent over $i \in [1\colon n]$. Thus,

$$\mathbb{E}\left[\prod_{i=1}^{n}\left\{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)\middle|\mathcal{H}^{(i)}_1, X^{(i)}_1\right] + c\right\}\right]$$

$$= \prod_{i=1}^{n}\mathbb{E}\left[\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T)\middle|\mathcal{H}^{(i)}_1, X^{(i)}_1\right] + c\right]$$

By law of iterated expectations,

$$= \prod_{i=1}^{n}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}^{(i)}_T) + c\right].$$

Thus we have shown that the desired result holds and all that is left is to show that Equation (72) holds.

Proof of Equation (72): The proof of Equation (72) leverages (i) the importance weights and (ii) conditional independence properties. Pick any $t \in [2:T]$ and let g be a function of $\mathcal{H}^{(i)}_t$. By law of iterated expectations,

$$\mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:t}\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right) + c\right\}\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:t}\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right) + c\right\}\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right]\right]$$

Note that the conditional expectation $\mathbb{E}\left[\prod_{i=1}^n\left\{\left(\hat{\pi}^{(i)}_{2:t}\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right) + c\right\}\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right]$ is only integrating over $\left\{A^{(i)}_t, R^{(i)}_t, X^{(i)}_{t+1}\right\}^n_{i=1}$. Additionally, note that conditional on $\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t$ that $\left\{A^{(i)}_t, R^{(i)}_t, X^{(i)}_{t+1}\right\}$ are independent over $i \in [1\colon n]$. Thus,

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:t}\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right)\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right] + c\right\}\right]$$

Since $\hat{\pi}^{(i)}_{2:t-1}$ is a constant given $\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t$,

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}^{(i)}_{2:t-1}\right)^{-1}\mathbb{E}\left[\left(\hat{\pi}^{(i)}_t\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right)\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right] + c\right\}\right]$$

Note $\mathbb{E}\left[\left(\hat{\pi}^{(i)}_t\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right)\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right] = \mathbb{E}\left[\left(\hat{\pi}^{(i)}_t\right)^{-1}\pi^{*,(i)}_t\left(\pi^{*,(i)}_t\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right)\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right]$
$= \mathbb{E}_{\pi^*_t}\left[\left(\pi^{*,(i)}_t\right)^{-1} g\left(\mathcal{H}^{(i)}_t, X^{(i)}_{t+1}\right)\middle|\mathcal{H}^{(1:n)}_{t-1}, X^{(1:n)}_t\right]$.

Also, note that the expectation $\mathbb{E}_{\pi_t^*}\left[\left(\pi_t^{*,(i)}\right)^{-1} g\left(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}\right) \middle| \mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right]$ integrates over $\left\{A_t^{(i)}, R_t^{(i)}, X_{t+1}^{(i)}\right\}$. Since actions are selected using π_t^* rather than $\hat{\pi}_t$, the distribution of $\left\{A_t^{(i)}, R_t^{(i)}, X_{t+1}^{(i)}\right\}$ depends only on $\mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}$ through $\mathcal{H}_{t-1}^{(i)}, X_t^{(i)}$. This means that $\mathbb{E}_{\pi_t^*}\left[\left(\pi_t^{*,(i)}\right)^{-1} g\left(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}\right) \middle| \mathcal{H}_{t-1}^{(1:n)}, X_t^{(1:n)}\right] = \mathbb{E}_{\pi_t^*}\left[\left(\pi_t^{*,(i)}\right)^{-1} g\left(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}\right) \middle| \mathcal{H}_{t-1}^{(i)}, X_t^{(i)}\right]$. Thus,

$$= \mathbb{E}\left[\prod_{i=1}^{n}\left\{\left(\hat{\pi}_{2:t-1}^{(i)}\right)^{-1}\mathbb{E}_{\pi_t^*}\left[\left(\pi_t^{*,(i)}\right)^{-1} g\left(\mathcal{H}_t^{(i)}, X_{t+1}^{(i)}\right) \middle| \mathcal{H}_{t-1}^{(i)}, X_t^{(i)}\right] + c\right\}\right].$$

We have now shown that Equation (72) holds. ■

D.2. Lemma 13: Maximal Inequality for Finite Class of Functions

Lemma 13 (Maximal Inequality for Finite Class of Functions) *We consider the problem setting as described in Section 1. We assume Condition 1. For any $\mathcal{F}$ that is a finite class of bounded, measurable functions of size $|\mathcal{F}| \geq 2$, for $\mathbb{G}_n(f) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\right]\right\}$,*

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\mathbb{G}_n(f)\right|\right] \leq C\Bigg\{\pi_{\min}^{-(T-1)}\max_{f\in\mathcal{F}}\frac{\|f\|_\infty}{\sqrt{n}}\log(|\mathcal{F}|) + \sqrt{\pi_{\min}^{-(T-1)}}\max_{f\in\mathcal{F}}\sqrt{\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}\sqrt{\log(|\mathcal{F}|)}\Bigg\} \quad (73)$$

for some universal positive constant C (specified in the proof).

Proof of Lemma 13 (Maximal Inequality for Finite Class of Functions) Our proof follows a very similar argument to Lemma 19.33 in Van der Vaart (2000). Specifically, our proof only deviates because we use our Lemma 11 to prove Equations (77) and (79) below.

Let u, v be non-negative, real-valued functions of $f \in \mathcal{F}$ such that

- $u(f) = 12\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]$
- $v(f) = 12\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}$

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\right] = \mathbb{E}\left[\max_{f\in\mathcal{F}}\left\{|\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|>u(f)/v(f)} + |\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|\leq u(f)/v(f)}\right\}\right]$$

$$\leq \mathbb{E}\left[\max_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|>u(f)/v(f)}\right] + \mathbb{E}\left[\max_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|\leq u(f)/v(f)}\right]$$

Let $\underline{\mathbb{G}_n}(f) \triangleq |\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|>u(f)/v(f)}$ and $\overline{\mathbb{G}_n}(f) \triangleq |\mathbb{G}_n(f)|\mathbb{I}_{|\mathbb{G}_n(f)|\leq u(f)/v(f)}$.

$$= \mathbb{E}\left[\max_{f\in\mathcal{F}}\underline{\mathbb{G}_n}(f)\right] + \mathbb{E}\left[\max_{f\in\mathcal{F}}\overline{\mathbb{G}_n}(f)\right]$$

$$\leq \mathbb{E}\left[\max_{f\in\mathcal{F}} \underline{\mathbb{G}_n}(f)/v(f)\right]\left(\max_{f\in\mathcal{F}} v(f)\right) + \mathbb{E}\left[\max_{f\in\mathcal{F}} \overline{\mathbb{G}_n}(f)/\sqrt{u(f)}\right]\left(\max_{f\in\mathcal{F}} \sqrt{u(f)}\right) \tag{74}$$

The main result we will show in this proof are the following

$$\mathbb{E}\left[\max_{f\in\mathcal{F}} \underline{\mathbb{G}_n}(f)/v(f)\right] \leq \log\left(1+|\mathcal{F}|\right) \tag{75}$$

$$\mathbb{E}\left[\max_{f\in\mathcal{F}} \overline{\mathbb{G}_n}(f)/\sqrt{u(f)}\right] \leq \sqrt{\log\left(1+|\mathcal{F}|\right)} \tag{76}$$

For now we take Equations (75) and (76) as given. Thus we have that Equation (74) can be upper bounded by the following:

$$\leq \log\left(1+|\mathcal{F}|\right)\left(\max_{f\in\mathcal{F}} \underbrace{12\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}}_{v(f)}\right) + \sqrt{\log\left(1+|\mathcal{F}|\right)}\left(\max_{f\in\mathcal{F}} \underbrace{\sqrt{12\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}}_{\sqrt{u(f)}}\right)$$

$$= 12\pi_{\min}^{-(T-1)}\left(\max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{n}}\right)\log\left(1+|\mathcal{F}|\right) + \sqrt{12\pi_{\min}^{-(T-1)}}\left(\max_{f\in\mathcal{F}} \sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}\right)\sqrt{\log\left(1+|\mathcal{F}|\right)}$$

Since $c_{\log} \triangleq \sup_{x\geq 2} \frac{\log(1+x)}{\log(x)}$ is bounded, $\frac{\log(1+|\mathcal{F}|)}{\log(|\mathcal{F}|)} \leq c_{\log}$ so, $1 \leq c_{\log}\frac{\log(|\mathcal{F}|)}{\log(1+|\mathcal{F}|)}$.

$$\leq c_{\log}12\pi_{\min}^{-(T-1)}\left(\max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{n}}\right)\log\left(|\mathcal{F}|\right) + \sqrt{c_{\log}12\pi_{\min}^{-(T-1)}}\left(\max_{f\in\mathcal{F}} \sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}\sqrt{\log\left(|\mathcal{F}|\right)}\right)$$

$$\leq 12\max\left(c_{\log}^{-1}, c_{\log}^{-1/2}\right)\left\{\pi_{\min}^{-(T-1)}\left(\max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{n}}\right)\log\left(|\mathcal{F}|\right) + \sqrt{\pi_{\min}^{-(T-1)}}\left(\max_{f\in\mathcal{F}} \sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]}\right)\sqrt{\log\left(|\mathcal{F}|\right)}\right\}.$$

The above implies that the desired result, Equation (73), holds. All that remains is to prove that Equations (75) and (76) hold.

Proving Equation (75) holds: Let $x > 0$. By Lemma 11, we have that

$$\mathbb{P}\left(|\underline{\mathbb{G}}_n(f)| > x\right) \leq 2\exp\left(-3\frac{x^2}{u(f) + xv(f)}\right) \leq 2\exp\left(-3\frac{x}{v(f)}\right), \tag{77}$$

where recall $u(f) = 12\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]$ and $v(f) = 12\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}$. The second inequality above holds since $u(f), v(f)$ are non-negative so $\frac{x^2}{u(f)+xv(f)} \leq \frac{x^2}{xv(f)} \leq \frac{x}{v(f)}$. We now show that the following is less than or equal to 1:

$$\mathbb{E}\left[e^{|\underline{\mathbb{G}}_n(f)|/v(f)}\right] - 1 = \mathbb{E}\left[\int_0^{|\underline{\mathbb{G}}_n(f)|/v(f)} e^x dx\right] = \mathbb{E}\left[\int_0^\infty \mathbb{I}_{x<|\underline{\mathbb{G}}_n(f)|/v(f)} e^x dx\right]$$

By Fubini's theorem, we can exchange the integrals,

$$= \int_0^\infty \mathbb{E}\left[\mathbb{I}_{x<|\underline{\mathbb{G}}_n(f)|/v(f)}\right] e^x dx = \int_0^\infty \mathbb{P}\Big(\left|\underline{\mathbb{G}}_n(f)\right| > xv(f)\Big) e^x dx$$

By Equation (77),

$$\leq 2\int_0^\infty e^{-3x}e^x dx = 2\int_0^\infty e^{-2x}dx = 2\left(\lim_{x\to\infty} -\frac{1}{2}e^{-2x} + \frac{1}{2}e^0\right) = 2\left(0 + \frac{1}{2}\right) = 1.$$

Thus we have that for $\gamma(x) = e^x - 1$,

$$\mathbb{E}\big[\gamma\big(\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\big)\big] = \mathbb{E}\big[\exp\big(\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\big)\big] - 1 \leq 1. \tag{78}$$

Note that $\gamma(x) = e^x - 1$ is convex and non-negative, so by Jensen's inequality,

$$\exp\left(\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\right]\right) - 1 = \gamma\left(\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\right]\right) \leq \mathbb{E}\left[\gamma\left(\max_{f\in\mathcal{F}}\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\right)\right]$$
$$\leq \sum_{f\in\mathcal{F}} \mathbb{E}\left[\gamma\left(\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\right)\right] \leq |\mathcal{F}|.$$

The last inequality above holds by Equation (78). By adding 1 and taking the log on both sides, we have Equation (75) holds, i.e., that

$$\mathbb{E}\left[\max_{f\in\mathcal{F}}\left|\underline{\mathbb{G}}_n(f)\right|/v(f)\right] \leq \log\left(|\mathcal{F}| + 1\right).$$

Proving Equation (76) holds: Let $x > 0$. By Lemma 11, we have that

$$\mathbb{P}\left(\left|\overline{\mathbb{G}}_n(f)\right| > x\right) \leq 2\exp\left(-3\frac{x^2}{u(f) + xv(f)}\right) \leq 2\exp\left(-3\frac{x^2}{u(f)}\right), \tag{79}$$

where recall $u(f) = 12\pi_{\min}^{-(T-1)}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})^2\right]$ and $v(f) = 12\pi_{\min}^{-(T-1)}\|f\|_\infty/\sqrt{n}$. The second inequality above holds since $u(f), v(f)$ are non-negative. We now show that the following is less than or equal to 1:

$$\mathbb{E}\left[e^{\left|\overline{\mathbb{G}_n}(f)\right|^2/u(f)}\right] - 1 = \mathbb{E}\left[\int_0^{\left|\overline{\mathbb{G}_n}(f)\right|^2/u(f)} e^x dx\right] = \mathbb{E}\left[\int_0^\infty \mathbb{I}_{x<\left|\overline{\mathbb{G}_n}(f)\right|^2/u(f)} e^x dx\right]$$

$$= \mathbb{E}\left[\int_0^\infty \mathbb{I}_{\sqrt{xu(f)}<\left|\overline{\mathbb{G}_n}(f)\right|} e^x dx\right]$$

By Fubini's theorem, we can exchange integrals,

$$= \int_0^\infty \mathbb{E}\left[\mathbb{I}_{\sqrt{xu(f)}<\left|\overline{\mathbb{G}_n}(f)\right|}\right] e^x dx = \int_0^\infty \mathbb{P}\left(\left|\overline{\mathbb{G}_n}(f)\right| > \sqrt{xu(f)}\right) e^x dx$$

By Equation (79),

$$\leq 2\int_0^\infty e^{-3x+x} dx = 2\int_0^\infty e^{-2x} dx = 2\left(\lim_{x\to\infty} -\frac{1}{2}e^{-2x} + \frac{1}{2}e^0\right) = 2\left(0+\frac{1}{2}\right) = 1.$$

Thus we have that for $\gamma_2(x) = e^{x^2} - 1$,

$$\gamma_2\left(\left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right) = \mathbb{E}\left[\exp\left(\left|\overline{\mathbb{G}_n}(f)\right|^2 / u(f)\right)\right] - 1 \leq 1. \tag{80}$$

Since $\gamma_2(x) = e^{x^2} - 1$ is convex, by Jensen's inequality,

$$\exp\left(\mathbb{E}\left[\max_{f\in\mathcal{F}} \left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right]^2\right) - 1 = \gamma_2\left(\mathbb{E}\left[\max_{f\in\mathcal{F}} \left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right]\right)$$

$$\leq \mathbb{E}\left[\gamma_2\left(\max_{f\in\mathcal{F}} \left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right)\right] \leq \sum_{f\in\mathcal{F}} \mathbb{E}\left[\gamma_2\left(\left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right)\right] \leq |\mathcal{F}|$$

The last inequality above holds by Equation (80). By adding 1, taking the log and the square-root on both sides, we have Equation (76) holds, i.e., that

$$\mathbb{E}\left[\max_{f\in\mathcal{F}} \left|\overline{\mathbb{G}_n}(f)\right| / \sqrt{u(f)}\right] \leq \sqrt{\log\left(|\mathcal{F}|+1\right)}. \quad \blacksquare$$

D.3. Lemma 14: Maximal Inequality as a Function of the Bracketing Integral

Lemma 14 (Maximal Inequality as a Function of the Bracketing Integral) *We consider the problem setting as described in Section 1. We assume Condition 1 holds. Let $\mathcal{F}$ be any class of real-valued measurable functions of $\mathcal{H}_T^{(i)}$ with $\mathbb{E}_{\pi^*_{2:T}}\left[(\pi^*_{2:T})^{-1} f(\mathcal{H}_T^{(i)})^2\right] \leq \delta^2$ for all $f \in \mathcal{F}$ and with a finite bracketing integral, i.e., for any $\eta > 0$, $\int_0^\eta \sqrt{\log N_{[\,]}\left(\epsilon, \mathcal{F}, L_2\left(\mathcal{P}_{\pi^*_{2:T}}\right)\right)} d\epsilon < \infty$. Additionally, we assume there exists an envelope function F, i.e., $|f(\mathcal{H}_T^{(i)})| < F(\mathcal{H}_T^{(i)}) < \infty$ w.p. 1. Then for $a(\delta) = \delta / \sqrt{\log N_{[\,]}\left(\delta, \mathcal{F}, L_2(\mathcal{P}_{\pi^*_{2:T}})\right)}$ and $\mathbb{G}_n(f) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} f(\mathcal{H}_T^{(i)})\right]\right\}$,*

$$\mathbb{E}^*\left[\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)|\right] \lesssim \int_0^\delta \sqrt{\log N_{[\,]}\left(\delta, \mathcal{F}, L_2\left(\mathcal{P}_{\pi^*_{2:T}}\right)\right)} d\epsilon$$
$$+ \sqrt{n}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} F(\mathcal{H}_T^{(i)}) \mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]. \tag{81}$$

Above $\lesssim$ means less than or equal to when scaled by universal positive constants.

Above $\mathbb{E}^*$ refers to outer expectations as defined in Van der Vaart (2000, Section 18.2).

Proof of Lemma 14 (Maximal Inequality as a Function of the Bracketing Integral) Our proof is almost identical to that of Van der Vaart (2000, Lemma 19.34) except that we use the maximal inequality in Lemma (14) instead of a maximal inequality for i.i.d. data; we include the full proof for clarity and completeness.

Note that by triangle inequality,

$$\mathbb{E}^*\left[\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|\right] \leq \mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\mathbb{G}_n\left(f\mathbb{I}_{F>\sqrt{n}a(\delta)}\right)\right|\right] + \mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\mathbb{G}_n\left(f\mathbb{I}_{F\leq\sqrt{n}a(\delta)}\right)\right|\right] \tag{82}$$

Bounding First Term in Equation (82): This term is to deal with potentially unbounded functions $f \in \mathcal{F}$.

$$\mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\mathbb{G}_n\left(f\mathbb{I}_{F>\sqrt{n}a(\delta)}\right)\right|\right]$$

By using the definition of $\mathbb{G}_n$,

$$= \mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)} - \mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]\right|\right]$$

By triangle inequality,

$$\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right|\right] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sup_{f\in\mathcal{F}}\left\{\left|\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} f(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]\right|\right\}$$

By Jensen's inequality,

$$\leq 2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1}\left|f(\mathcal{H}_T^{(i)})\right|\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]$$

By our envelope function F,

$$\leq 2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\left[\left(\hat{\pi}_{2:T}^{(i)}\right)^{-1} F(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]$$

$$= 2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} F(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]$$

Since the expectation above is indexed by the deterministic policy $\pi_{2:T}^*$, $\mathcal{H}_T^{(i)}$ within the expectation are i.i.d.

$$= 2\sqrt{n}\mathbb{E}_{\pi_{2:T}^*}\left[\left(\pi_{2:T}^{*,(i)}\right)^{-1} F(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})>\sqrt{n}a(\delta)}\right]$$

This us gives us the second part of bound (81).

Bounding Second Term in Equation (82): Thus, we focus on bounding the following:

$$\mathbb{E}^*\left[\sup_{f\in\mathcal{F}}\left|\mathbb{G}_n f\mathbb{I}_{F\leq\sqrt{n}a(\delta)}\right|\right]$$

We now deal with the class of functions $\bar{\mathcal{F}} := \left\{f\mathbb{I}_{F\leq\sqrt{n}a(\delta)} : f\in\mathcal{F}\right\}$. We first show that $N_{[\,]}\big(\epsilon,\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\leq N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$.

- By definition of bracketing numbers, we can cover $\mathcal{F}$ with $N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$ brackets, each with size at most ϵ. Specifically, we can find brackets $[l_j,u_j]$ for $j\in\left[1\colon N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\right]$ that cover $\mathcal{F}$ and such that $\mathbb{E}_{\pi^*_{2:T}}\left[(u_j-l_j)^2\right]\leq\epsilon$ for all brackets $[l_j,u_j]$.
- Note that brackets $\left[l_j\mathbb{I}_{F\leq\sqrt{n}a(\delta)},\ u_j\mathbb{I}_{F\leq\sqrt{n}a(\delta)}\right]$ for $j\in\left[1\colon N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\right]$ cover $\bar{\mathcal{F}}$.
- Additionally, note that $\mathbb{E}_{\pi^*_{2:T}}\left[\Big([u_j-l_j]\mathbb{I}_{F\leq\sqrt{n}a(\delta)}\Big)^2\right]\leq\mathbb{E}_{\pi^*_{2:T}}\left[(u_j-l_j)^2\right]\leq\epsilon$.

Thus, we have that

$$N_{[\,]}\big(\epsilon,\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\leq N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big). \tag{83}$$

Desiderata for Nested Partitions We now assume the existence of nested partitions of $\bar{\mathcal{F}}$ that satisfy certain conditions. We will finish the proof assuming these partitions exist and conclude by constructing these partitions.

High level, we assume we have nested partitions of $\bar{\mathcal{F}}$ that are indexed by positive integers q. These partitions are designed to become increasingly fine-grained as q increases. Specifically the "size" of each piece of the partition will be on the order of 2^{-q}, i.e., the "size" of the partitions will halve as q increases by 1. The partitions are nested in that each partition piece at level $q+1$ is a subset of some partition piece at level q.

We pick q_0 to be a positive integer such that $\delta<2^{-(q_0+2)}\leq 2\delta$. For every integer $q\geq q_0$ we have a partition of $\bar{\mathcal{F}}$, which we write as $\{\bar{\mathcal{F}}_{q,j}\}_{j=1}^{N_q}$; we assume that $N_{q_0}=N_{[\,]}\big(2^{-q_0},\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$. These partitions are nested in that for each $q\geq q_0+1$ and for every $j\in[1\colon N_q]$, we have that the partition piece $\bar{\mathcal{F}}_{q,j}$ is a subset of some partition piece $\bar{\mathcal{F}}_{q-1,k}$ for some $k\in[1\colon N_{q-1}]$. Moreover, we further assume the following:

- **Requirement on the "size" of partition pieces:** For each partition q and partition piece $j\in[1\colon N_q]$, let $\Delta_{q,j}$ be a measurable function of $\mathcal{H}_T^{(i)}$ such that $\sup_{f,g\in\bar{\mathcal{F}}_{q,j}}|f-g|\leq\Delta_{q,j}$ and
$$\mathbb{E}_{\pi^*_{2:T}}\left[(\hat{\pi}^{(i)}_{2:T})^{-1}\Delta_{q,j}(\mathcal{H}_T^{(i)})^2\right]\leq 2^{-2q}. \tag{84}$$
- **Requirement on how the number of partition pieces grows as the "size" goes to zero:**
$$\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}\lesssim\int_0^{\delta}\sqrt{\log N_{[\,]}\big(\epsilon,\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon \tag{85}$$

Main Argument Assuming Desired Nested Partitions Now assuming such partitions described above exist, we continue with the argument.

For every partition piece $\bar{\mathcal{F}}_{q,j}$, we choose a arbitrary point $\bar{f}_{q,j}$ in that partition piece, i.e., for each $q \geq q_0$ and every $j \in [1\colon N_q]$ we choose a point $\bar{f}_{q,j} \in \mathcal{F}_{q,j}$. We also define functions $\lambda_q : \bar{\mathcal{F}} \mapsto \bar{\mathcal{F}}$ that maps each function $\bar{f} \in \mathcal{F}$ to these points $\{\bar{f}_{q,j}\}_{j=1}^{N_q}$; specifically, for any $\bar{f} \in \mathcal{F}$ we can find some partition piece $\mathcal{F}_{q,j}$ such that $\bar{f} \in \mathcal{F}_{q,j}$ and we map that $\bar{f}$ to the point $\bar{f}_{q,j}$.

Note that for any integer $Q > q_0$, by telescoping series, for any $\bar{f} \in \bar{\mathcal{F}}$,

$$\bar{f}(\mathcal{H}_T^{(i)}) = \lambda_{q_0}\bar{f}(\mathcal{H}_T^{(i)}) + \sum_{q=q_0}^{Q}\left\{\lambda_{q+1}\bar{f}(\mathcal{H}_T^{(i)}) - \lambda_q\bar{f}(\mathcal{H}_T^{(i)})\right\} + \bar{f}(\mathcal{H}_T^{(i)}) - \lambda_{Q+1}\bar{f}(\mathcal{H}_T^{(i)})$$

$$= \lambda_{q_0}\bar{f}(\mathcal{H}_T^{(i)}) + \sum_{q=q_0}^{\infty}\mathbb{I}_{q\leq Q}\left\{\lambda_{q+1}\bar{f}(\mathcal{H}_T^{(i)}) - \lambda_q\bar{f}(\mathcal{H}_T^{(i)})\right\} + \sum_{q=q_0}^{\infty}\mathbb{I}_{q=Q+1}\left\{\bar{f}(\mathcal{H}_T^{(i)}) - \lambda_q\bar{f}(\mathcal{H}_T^{(i)})\right\}. \tag{86}$$

For any $\bar{f} \in \bar{\mathcal{F}}$, we define $Q_{\bar{f}}(\mathcal{H}_T^{(i)}) \in [q_0, \infty]$ to be a random variable representing the maximum partition level with no bound violations up to that level. Specifically,

$$Q_{\bar{f}}(\mathcal{H}_T^{(i)}) \triangleq \left\{\max_{q\geq q_0} \text{ s.t. } \sum_{j=1}^{N_q}\mathbb{I}_{\bar{f}\in\bar{\mathcal{F}}_{q,j}}\Delta_{p,j}(\mathcal{H}_T^{(i)}) \leq \sqrt{n}a_p \text{ for all } p \in [q_0 : q]\right\}$$

where $a_q = 2^{-q}/\sqrt{\log N_{q+1}}$. Thus, by replacing Q with $Q_{\bar{f}}$ and by applying $\mathbb{G}_n$ to both sides of Equation (86),

$$\mathbb{G}_n(\bar{f}) = \mathbb{G}_n(\lambda_{q_0}\bar{f}) + \sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q\leq Q_{\bar{f}}}\left(\lambda_{q+1}\bar{f} - \lambda_q\bar{f}\right)\right) + \sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar{f}}+1}\left(\bar{f} - \lambda_q\bar{f}\right)\right)$$

Thus, we have that by triangle inequality

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n(\bar{f})\right|\right] \leq \underbrace{\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n(\lambda_{q_0}\bar{f})\right|\right]}_{(i)}$$
$$+ \underbrace{\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q\leq Q_{\bar{f}}}\left(\lambda_{q+1}\bar{f} - \lambda_q\bar{f}\right)\right)\right|\right]}_{(ii)}$$
$$+ \underbrace{\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar{f}}+1}\left(\bar{f} - \lambda_q\bar{f}\right)\right)\right|\right]}_{(iii)}. \tag{87}$$

Below we will show the following results:

- Bounding term (i)

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n(\lambda_{q_0}\bar{f})\right|\right] \leq 2\pi_{\min}^{-(T-1)}2^{-q_0}\sqrt{\log(N_{q_0})} \tag{88}$$

- Bounding term (ii)

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\big(\mathbb{I}_{q\le Q_{\bar{f}}}\left(\lambda_{q+1}\bar{f}-\lambda_q\bar{f}\right)\big)\right|\right]\lesssim 2\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q} \tag{89}$$

- Bounding term (iii)

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar{f}}+1}\left(\bar{f}-\lambda_q\bar{f}\right)\right)\right|\right]\lesssim 6\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}. \tag{90}$$

For now we assume the above three equations hold. Thus, we can upper bound Equation (87) as follows:

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n(\bar{f})\right|\right]\lesssim 2\pi_{\min}^{-(T-1)}2^{-q_0}\sqrt{\log N_{q_0}}+8\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}$$

$$\le 10\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}$$

By Equation (85),

$$\lesssim \pi_{\min}^{-(T-1)}\int_0^{\delta}\sqrt{\log\big(N_{[\,]}\big(\epsilon,\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon\le \pi_{\min}^{-(T-1)}\int_0^{\delta}\sqrt{\log\big(N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon.$$

The last inequality above holds by Equation (83). We now show that Equations (88), (89), and (90) hold.

Equation (88): Bounding term (i)

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n(\lambda_{q_0}\bar{f})\right|\right]=\mathbb{E}\left[\max_{j\in[1:\,N_{q_0}]}\left|\mathbb{G}_n(\bar{f}_{q_0,j})\right|\right]$$

By Lemma 13, a maximal inequality for finite classes of functions,

$$\lesssim \pi_{\min}^{-(T-1)}\max_{j\in[1:\,N_{q_0}]}\frac{\|\bar{f}_{q_0,j}\|_\infty}{\sqrt{n}}\log N_{q_0}$$
$$+\sqrt{\pi_{\min}^{-(T-1)}}\max_{j\in[1:\,N_{q_0}]}\sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1}\bar{f}_{q_0,j}(\mathcal{H}_T^{(i)})^2\right]}\sqrt{\log N_{q_0}}. \tag{91}$$

- Note that since $\bar{f}(\mathcal{H}_T^{(i)})=f(\mathcal{H}_T^{(i)})\mathbb{I}_{F(\mathcal{H}_T^{(i)})\le\sqrt{n}a(\delta)}\le\sqrt{n}a(\delta)$ w.p. 1, we get the first inequality below:

$$\max_{j\in[1:\,N_{q_0}]}\left\{\|\bar{f}_{q_0,j}\|_\infty\right\}\le\sqrt{n}a(\delta)\le\sqrt{n}a_{q_0}.$$

  For the second inequality above, recall $a(\delta)\triangleq\delta/\sqrt{\log N_{[\,]}\big(\delta,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}$ and $a_{q_0}\triangleq 2^{-q_0}/\sqrt{\log N_{q_0+1}}$. Since $\delta<2^{-(q_0+2)}\le 2\delta$, by Equation (83), $N_{[\,]}\big(\delta,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\ge N_{[\,]}\big(2^{-(q_0+2)},\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\ge N_{[\,]}\big(2^{-(q_0+1)},\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)\ge N_{[\,]}\big(2^{-(q_0+1)},\bar{\mathcal{F}},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$ $=N_{q_0+1}$. So, $a(\delta)\le 2^{-(q_0+2)}/\sqrt{\log N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}\le 2^{-q_0}/\sqrt{\log N_{[\,]}\big(\epsilon,\mathcal{F},L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}$ $\le 2^{-q_0}/\sqrt{\log N_{q_0+1}}=a_{q_0}$.

- $\max_{j\in[1\colon N_q]} \sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1}\bar{f}_{q_0,j}(\mathcal{H}^{(i)}_T)^2\right]} \le \delta \le 2^{-(q_0+2)}$ since we choose q_0 such that $\delta < 2^{-(q_0+2)} \le 2\delta$.

Thus, we can upper bound Equation (91) as follows:

$$\le \pi_{\min}^{-(T-1)} a_{q_0} \log N_{q_0} + \sqrt{\pi_{\min}^{-(T-1)}}\sqrt{2^{-2(q_0+2)}}\sqrt{\log N_{q_0}}$$

Since $a_{q_0} \triangleq 2^{-q_0}/\sqrt{\log N_{q_0+1}}$ and $N_{q_0+1} \ge N_{q_0}$,

$$\le \pi_{\min}^{-(T-1)} 2^{-q_0}\sqrt{\log N_{q_0}} + \sqrt{\pi_{\min}^{-(T-1)}} 2^{-(q_0+2)}\sqrt{\log N_{q_0}} \le 2\pi_{\min}^{-(T-1)} 2^{-q_0}\sqrt{\log N_{q_0}}.$$

Thus, we have that Equation (88) holds.

Bounding term (ii):
By triangle inequality,

$$\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q\le Q_{\bar{f}}}\left(\lambda_{q+1}\bar{f}-\lambda_q\bar{f}\right)\right)\right|\right] \le \sum_{q=q_0}^{\infty}\mathbb{E}^*\left[\sup_{\bar{f}\in\mathcal{F}}\left|\mathbb{G}_n\left(\mathbb{I}_{q\le Q_{\bar{f}}}\left(\lambda_{q+1}\bar{f}-\lambda_q\bar{f}\right)\right)\right|\right]$$

Since $\mathbb{I}_{q\le Q_{\bar{f}}} = \mathbb{I}_{q\le Q_{\lambda_q\bar{f}}}$,

$$= \sum_{q=q_0}^{\infty}\mathbb{E}\left[\max_{j\in[1\colon N_q]}\left|\mathbb{G}_n\left(\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}\left(\bar{f}_{q+1,j}-\bar{f}_{q,j}\right)\right)\right|\right]$$

By Lemma 13, a maximal inequality for finite classes of functions,

$$\lesssim \sum_{q=q_0}^{\infty}\pi_{\min}^{-(T-1)}\max_{j\in[1\colon N_{q_0}]}\frac{\left\|\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}\left(\bar{f}_{q+1,j}-\bar{f}_{q,j}\right)\right\|_\infty}{\sqrt{n}}\log N_q$$
$$+\sum_{q=q_0}^{\infty}\sqrt{\pi_{\min}^{-(T-1)}}\max_{j\in[1\colon N_q]}\sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1}\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}\left(\bar{f}_{q+1,j}-\bar{f}_{q,j}\right)^2\right]}\sqrt{\log N_q}. \quad (92)$$

- Note that by the definition of $\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}$ and by our nested partitions, we have that $\left\|\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}\left(\bar{f}_{q+1,j}-\bar{f}_{q,j}\right)\right\|_\infty \le \sup_{\bar{f},\bar{f}'\in\bar{\mathcal{F}}_{q,j}}\left\|\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}|f-f'|\right\|_\infty \le \left\|\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right\|_\infty \le \sqrt{n}a_q$.
- By our nested partitions, we have that $\lambda_{q+1}\bar{f}, \lambda_q\bar{f}$ are in the same q^{th}-level partition piece, i.e., $\lambda_{q+1}\bar{f}, \lambda_q\bar{f} \in \bar{\mathcal{F}}_{q,j}$ for some $\bar{\mathcal{F}}_{q,j}$ with $j \in [1\colon N_q]$. Thus,

$$\mathbb{E}_{\pi^*_{2:T}}\left[\mathbb{I}_{q\le Q_{\bar{f}_{q,j}}(\mathcal{H}^{(i)}_T)}(\pi^{*,(i)}_{2:T})^{-1}\left(\bar{f}_{q+1,j}(\mathcal{H}^{(i)}_T)-\bar{f}_{q,j}(\mathcal{H}^{(i)}_T)\right)^2\right]$$
$$\le \sup_{\bar{f},\bar{f}'\in\bar{\mathcal{F}}_{q,j}}\mathbb{E}_{\pi^*_{2:T}}\left[(\pi^{*,(i)}_{2:T})^{-1}\left(\bar{f}(\mathcal{H}^{(i)}_T)-\bar{f}'(\mathcal{H}^{(i)}_T)\right)^2\right]$$
$$\le \mathbb{E}_{\pi^*_{2:T}}\left[(\pi^{*,(i)}_{2:T})^{-1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)^2\right].$$

Moreover, by properties of our partitions,

$$\max_{j\in[1:\,N_q]}\sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[\mathbb{I}_{q\le Q_{\bar f_{q,j}}(\mathcal{H}_T^{(i)})}(\pi^{*,(i)}_{2:T})^{-1}\left(\bar f_{q+1,j}(\mathcal{H}_T^{(i)})-\bar f_{q,j}(\mathcal{H}_T^{(i)})\right)^2\right]} \le \max_{j\in[1:\,N_q]}\sqrt{\mathbb{E}_{\pi^*_{2:T}}\left[(\pi^{*,(i)}_{2:T})^{-1}\Delta_{q,j}(\mathcal{H}_T^{(i)})^2\right]} \le \sqrt{2^{-2q}}.$$

The last inequality above holds by the size property of our partitions.

We have that Equation (92) is upper bounded by the following:

$$\le \sum_{q=q_0}^{\infty}\left\{\pi_{\min}^{-(T-1)}a_q\log N_q+\sqrt{\pi_{\min}^{-(T-1)}}2^{-q}\sqrt{\log N_q}\right\} \le 2\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}.$$

The last inequality above holds because $a_q \triangleq 2^{-q}/\sqrt{\log N_{q+1}}$ and $N_{q+1}\ge N_q$. Thus, we have that Equation (89) holds.

Bounding term (iii):
By triangle inequality,

$$\mathbb{E}^*\left[\sup_{\bar f\in\mathcal{F}}\left|\sum_{q=q_0}^{\infty}\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar f}+1}\left(\bar f-\lambda_q\bar f\right)\right)\right|\right] \le \sum_{q=q_0}^{\infty}\mathbb{E}^*\left[\sup_{\bar f\in\mathcal{F}}\left|\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar f}+1}\left(\bar f-\lambda_q\bar f\right)\right)\right|\right] \quad (93)$$

Note that for some functions f, g such that $\left|f(\mathcal{H}_T^{(i)})\right| \le g(\mathcal{H}_T^{(i)})$,

$$\begin{aligned}|\mathbb{G}_n(f)| &\le \frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left(\hat\pi^{(i)}_{2:T}\right)^{-1}f(\mathcal{H}_T^{(i)})\right|+\sqrt{n}\left|\mathbb{E}\left[\left(\hat\pi^{(i)}_{2:T}\right)^{-1}f(\mathcal{H}_T^{(i)})\right]\right| \\ &\le \frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left(\hat\pi^{(i)}_{2:T}\right)^{-1}g(\mathcal{H}_T^{(i)})\right|+\sqrt{n}\left|\mathbb{E}\left[\left(\hat\pi^{(i)}_{2:T}\right)^{-1}g(\mathcal{H}_T^{(i)})\right]\right| \\ &\le |\mathbb{G}_n(g)|+2\sqrt{n}\left|\mathbb{E}\left[\left(\hat\pi^{(i)}_{2:T}\right)^{-1}g(\mathcal{H}_T^{(i)})\right]\right|.\end{aligned}$$

Note that $\left|\mathbb{I}_{q=Q_{\bar f}(\mathcal{H}_T^{(i)})+1}\left(\bar f(\mathcal{H}_T^{(i)})-\lambda_q\bar f(\mathcal{H}_T^{(i)})\right)\right| \le \mathbb{I}_{q=Q_{\bar f}(\mathcal{H}_T^{(i)})+1}\sum_{j=1}^{N_q}\mathbb{I}_{\bar f\in\mathcal{F}_{q,j}}\Delta_{q,j}(\mathcal{H}_T^{(i)})$. So we can upper bound Equation (93) as follows:

$$\le \sum_{q=q_0}^{\infty}\mathbb{E}^*\left[\sup_{\bar f\in\mathcal{F}}\left|\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar f}+1}\sum_{j=1}^{N_q}\mathbb{I}_{\bar f\in\mathcal{F}_{q,j}}\Delta_{q,j}\right)\right|\right] + 2\sqrt{n}\sum_{q=q_0}^{\infty}\sup_{\bar f\in\mathcal{F}}\left|\mathbb{E}\left[\left(\hat\pi^{(i)}_{2:T}\right)^{-1}\mathbb{I}_{q=Q_{\bar f}(\mathcal{H}_T^{(i)})+1}\sum_{j=1}^{N_q}\mathbb{I}_{\bar f\in\mathcal{F}_{q,j}}\Delta_{q,j}(\mathcal{H}_T^{(i)})\right]\right|$$

Since $\mathbb{I}_{q\le Q_{\bar{f}}+1} = \mathbb{I}_{q\le Q_{\lambda_q \bar{f}}+1}$,

$$= \sum_{q=q_0}^{\infty} \mathbb{E}\left[\max_{j\in[1:\,N_q]} \left|\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar{f}_{q,j}}+1}\,\Delta_{q,j}\right)\right|\right] + 2\sqrt{n}\sum_{q=q_0}^{\infty}\max_{j\in[1:\,N_q]}\mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1}\mathbb{I}_{q=Q_{\bar{f}_{q,j}}(\mathcal{H}^{(i)}_T)+1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right] \quad (94)$$

- Note that by our nested partitions and by the definition of $Q_{\bar{f}_{q,j}}$, we have that $\left\|\mathbb{I}_{q=Q_{\bar{f}_{q,j}}(\mathcal{H}^{(i)}_T)+1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right\|_\infty > \sqrt{n}a_q$, so $\left\|\mathbb{I}_{q=Q_{\bar{f}_{q,j}}(\mathcal{H}^{(i)}_T)+1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right\|_\infty(\sqrt{n}a_q)^{-1} > 1$.

  $$\begin{aligned}
  &2\sqrt{n}\max_{j\in[1:\,N_q]}\left|\mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1}\mathbb{I}_{q=Q_{\bar{f}_{q,j}}(\mathcal{H}^{(i)}_T)+1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right]\right| \\
  &\le \frac{2\sqrt{n}}{\sqrt{n}a_q}\max_{j\in[1:\,N_q]}\mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)^2\right] = \frac{2}{a_q}\max_{j\in[1:\,N_q]}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)^2\right] \\
  &= \frac{2}{a_q}\max_{j\in[1:\,N_q]}\mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)^2\right] \le \frac{2\cdot 2^{-2q}}{a_q}.
  \end{aligned}$$

  The last inequality above holds by the size property of our partitions.

- By our definition of $Q_{\bar{f}_{q,j}}$ and our nested partitions, we have that $\mathbb{E}_{\pi^*_{2:T}}\left[(\pi^{*,(i)}_{2:T})^{-1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)^2\right] \le 2^{-2q}$. By our definition of $\mathbb{I}_{q=Q_{\bar{f}_{q,j}}+1}$ and by our nested partitions, we have that $\left\|\mathbb{I}_{q=Q_{\bar{f}_{q,j}}+1}\Delta_{q,j}(\mathcal{H}^{(i)}_T)\right\|_\infty \le \sqrt{n}a_{q-1}$. Thus, by Lemma 13,

  $$\sum_{q=q_0}^{\infty}\mathbb{E}^*\left[\max_{j\in[1:\,N_q]}\left|\mathbb{G}_n\left(\mathbb{I}_{q=Q_{\bar{f}_{q,j}}+1}\,\Delta_{q,j}\right)\right|\right] \lesssim \sum_{q=q_0}^{\infty}\pi_{\min}^{-(T-1)}a_{q-1}\log N_q + \sqrt{\pi_{\min}^{-(T-1)}}2^{-q}\sqrt{\log N_q}.$$

The above observations allow us to upper bound Equation (94) as follows:

$$\lesssim \sum_{q=q_0}^{\infty}\left\{\pi_{\min}^{-(T-1)}a_{q-1}\log N_q + \sqrt{\pi_{\min}^{-(T-1)}}2^{-q}\sqrt{\log N_q}\right\} + 2\sum_{q=q_0}^{\infty}\frac{2^{-2q}}{a_q}$$

Since $a_q \triangleq 2^{-q}/\sqrt{\log N_{q+1}}$ and $N_{q+1} \ge N_q$,

$$= \sum_{q=q_0}^{\infty}\left\{\pi_{\min}^{-(T-1)}2^{-(q-1)}\sqrt{\log N_q} + \sqrt{\pi_{\min}^{-(T-1)}}2^{-q}\sqrt{\log N_q}\right\} + 2\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_{q+1}}$$

Note that $2\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_{q+1}} = 4\sum_{q=q_0}^{\infty}2^{-(q+1)}\sqrt{\log N_{q+1}} = 4\sum_{q=q_0+1}^{\infty}2^{-q}\sqrt{\log N_q}$.

$$\le 6\pi_{\min}^{-(T-1)}\sum_{q=q_0}^{\infty}2^{-q}\sqrt{\log N_q}.$$

Thus, we have that Equation (90) holds.

Construct nested partitions: We now construct nested partitions that satisfy the conditions described previously, particularly Equations (84) and (85).

By our bracketing number assumption, for every integer $q \geq q_0$, we can find $N_q^* \triangleq N_{[\,]}\big(2^{-q}, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$ bracketing functions $\big\{\big[l_{q,j}, u_{q,j}\big]\big\}_{j=1}^{N_q^*}$ of size at most 2^{-q} that cover $\bar{\mathcal{F}}$. These brackets form a partition of $\bar{\mathcal{F}}$, which we write as $\big\{\bar{\mathcal{F}}^*_{q,j}\big\}_{j=1}^{N_q^*}$. Note that these partitions are not necessarily nested.

We take intersections of these partitions to construct a set of nested partitions $\big\{\bar{\mathcal{F}}_{q,j}\big\}_{j=1}^{N_q}$ for all integers $q \geq q_0$.

- For partition $\big\{\bar{\mathcal{F}}_{q_0,j}\big\}_{j=1}^{N_{q_0}}$, we simply set $\bar{\mathcal{F}}_{q_0,j} \triangleq \bar{\mathcal{F}}^*_{q_0,j}$ for all $j \in \big[1\colon N^*_{q_0}\big]$. This means that $N_{q_0} \triangleq N^*_{q_0}$.
- For partition $\big\{\bar{\mathcal{F}}_{q_0+1,j}\big\}_{j=1}^{N_{q_0+1}}$, we set partition pieces $\bar{\mathcal{F}}_{q_0+1,j}$ for all $j \in \big[1\colon N^*_{q_0}\big]$ to be the intersections between all pairs of partition pieces $\bar{\mathcal{F}}^*_{q_0,k}$ and $\bar{\mathcal{F}}^*_{q_0+1,l}$ for $k \in \big[1\colon N_{q_0}\big]$ and $l \in \big[1\colon N_{q_0}+1\big]$. This means that $N_{q_0+1} \triangleq N^*_{q_0} \cdot N^*_{q_1}$. Note that it could be that some partition pieces $\bar{\mathcal{F}}_{q_0+1,j}$ are empty, e.g., if the original partitions $\big\{\bar{\mathcal{F}}^*_{q,j}\big\}_{j=1}^{N_q^*}$ were already nested for $q = q_0, q_0+1$.
- For general $q \geq q_0$, we set partition pieces $\bar{\mathcal{F}}_{q,j}$ for $j \in \big[1\colon N_q^*\big]$ to be the intersections between all possible combinations in which we take one partition piece from $\big\{\bar{\mathcal{F}}^*_{p,j}\big\}_{j=1}^{N_p^*}$ for each $p \in [q_0\colon q]$. This means that each partition pieces $\bar{\mathcal{F}}_{q,j}$ is the intersection between $\bar{\mathcal{F}}^*_{q_0,k_{q_0}}, \bar{\mathcal{F}}^*_{q_0+1,k_{q_0+1}}, ..., \bar{\mathcal{F}}^*_{q_0+1,k_q}$ for $k_{q_0} \in \big[1\colon N_{q_0}\big], k_{q_0+1} \in \big[1\colon N_{q_0}+1\big], ..., k_q \in \big[1\colon N_q\big]$. This means that there are $N_q \triangleq \prod_{p=q_0}^{q} N_p^*$ total partition pieces.

Recall that the partitions were defined by bracketing functions $\big\{\big[l_{q,j}, u_{q,j}\big]\big\}_{j=1}^{N_q^*}$. This means that $\bar{\mathcal{F}}_{q,j}$ for $j \in \big[1\colon N_q\big]$ is covered by the bracket $\big[l_{q,j}, u_{q,j}\big]$. Moreover, $\mathbb{E}_{\pi^*_{2:T}}\big[\big(u_{q,j}(H_T^{(i)}) - l_{q,j}(H_T^{(i)})\big)^2\big] \leq 2^{-q}$. Thus, we define $\Delta_{q,j} \triangleq u_{q,j} - l_{q,j}$; note that this choice of $\Delta_{q,j}$ satisfies the conditions of Equation (84).

We now show that Equation (85) holds, i.e., the number of sets in the partition grows at a bounded rate as the size of the partition pieces goes to zero:

$$\sum_{q=q_0}^{\infty} 2^{-q}\sqrt{\log N_q} = \sum_{q=q_0}^{\infty} 2^{-q}\sqrt{\log\Big(\prod_{p=q_0}^{q} N_p^*\Big)} = \sum_{q=q_0}^{\infty} 2^{-q}\sqrt{\sum_{p=q_0}^{q}\log N_p^*}$$

Note that $\sqrt{\sum_{p=q_0}^{q}\log N_p^*} \leq \sum_{p=q_0}^{q}\sqrt{\log N_p^*}$ because $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for any positive non-negative values a, b.

$$\leq \sum_{q=q_0}^{\infty} 2^{-q}\sum_{p=q_0}^{q}\sqrt{\log N_p^*} = \sum_{q=q_0}^{\infty} 2^{-q}\sum_{p=q_0}^{\infty}\mathbb{I}_{p\leq q}\sqrt{\log N_p^*} = \sum_{p=q_0}^{\infty}\sqrt{\log N_p^*}\sum_{q=q_0}^{\infty} 2^{-q}\mathbb{I}_{p\leq q}$$

For the last equality above, we can exchange the infinite summations above by Fubini's theorem because below we will show that the last term above is bounded.

Since $\sum_{q=q_0}^{\infty} 2^{-q}\mathbb{I}_{p\leq q} = \sum_{q=p}^{\infty} 2^{-q} = 2^{-(p-1)}$,

$$= \sum_{p=q_0}^{\infty} 2^{-(p-1)}\sqrt{\log N_p^*} = 4\sum_{p=q_0}^{\infty} 2^{-(p+1)}\sqrt{\log N_p^*} = 4\sum_{p=q_0}^{\infty} 2^{-(p+1)}\sqrt{\log N_{[\,]}\big(2^{-p}, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}$$

Since $N_{[\,]}\big(2^{-p}, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)$ is monotonically increasing as p increases by lower Darboux sums, we have the following upper bound:

$$\leq 4\int_0^{2^{-q_0}} \sqrt{\log N_{[\,]}\big(\epsilon, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon$$

Since we chose q_0 such that $\delta < 2^{-(q_0+2)} \leq 2\delta$,

$$\leq 4\int_0^{8\delta} \sqrt{\log N_{[\,]}\big(\epsilon, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon \leq 32\int_0^{\delta} \sqrt{\log N_{[\,]}\big(\epsilon, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon < \infty.$$

The second to last inequality above holds because $N_{[\,]}\big(\epsilon, \bar{\mathcal{F}}, L_2(\mathcal{P}_{\pi^*_{2:T}})$ is monotonically increasing as ϵ goes to zero. The last inequality above holds by our finite bracketing integral assumption. ■

D.4. Theorem 15: Functional Asymptotic Normality under Finite Bracketing Integral

Theorem 15 (Functional Asymptotic Normality under Finite Bracketing Integral) *We consider the problem setting as described in Section 1. We assume Conditions 1 and 2 and that $\hat{\theta}_t - \theta_t^* = O_P(1/\sqrt{n})$ for all $t \in [1\colon T-1]$. Let $\mathcal{F}$ be any class of real-valued measurable functions f of $\mathcal{H}_T^{(i)}$ such that for some $\alpha > 0$, $\mathbb{E}_{\pi^*_{2:T}}\left[f\big(\mathcal{H}_T^{(i)}\big)^{4+\alpha}\right] < \infty$ and $\int_0^1 \sqrt{\log N_{[\,]}\big(\epsilon, \mathcal{F}, L_2(\mathcal{P}_{\pi^*_{2:T}})\big)}d\epsilon < \infty$. Then, for $\mathbb{G}_n(f) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1} f(\mathcal{H}_T^{(i)}) - \mathbb{E}\big[\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1} f(\mathcal{H}_T^{(i)})\big]\right\}$, the empirical process $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ converges in distribution to $\mathbb{G}_{\mathcal{F}}$ a mean-zero Gaussian process in $l^\infty(\mathcal{F})$ (the collection of all bounded functions from $\mathcal{F}$ to $\mathbb{R}$) with the following covariance function:*

$$\mathbb{E}\big[\mathbb{G}_{\mathcal{F}}(f)\mathbb{G}_{\mathcal{F}}(g)\big] \triangleq \mathbb{E}_{\pi^*_{2:T}}\left[\big(\pi^{*,(i)}_{2:T}\big)^{-2} f(\mathcal{H}_T^{(i)})g(\mathcal{H}_T^{(i)})\right] - \mathbb{E}_{\pi^*_{2:T}}\left[\big(\pi^{*,(i)}_{2:T}\big)^{-1} f(\mathcal{H}_T^{(i)})\right]\mathbb{E}_{\pi^*_{2:T}}\left[\big(\pi^{*,(i)}_{2:T}\big)^{-1} g(\mathcal{H}_T^{(i)})\right].$$

Proof of Theorem 15 By Van der Vaart (2000, Theorem 18.14), to show the desired result it is sufficient to show that the following two properties hold:

1. **Joint Convergence of Marginals** For any finite number of functions $f_1, f_2, ..., f_K \in \mathcal{F}$,

$$(\mathbb{G}_n(f_1), \mathbb{G}_n(f_2), ..., \mathbb{G}_n(f_K)) \xrightarrow{D} (\mathbb{G}_{\mathcal{F}}(f_1), \mathbb{G}_{\mathcal{F}}(f_2), ..., \mathbb{G}_{\mathcal{F}}(f_K))$$

2. **Asymptotically Tight** For any $\epsilon, \eta > 0$, there exists a partition of $\mathcal{F}$ into finitely many sets $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_j$ such that

$$\limsup_{n\to\infty} \mathbb{P}^*\left(\sup_{i\in[1\colon n]} \sup_{f,f'\in\mathcal{F}_i} \big|\mathbb{G}_n(f) - \mathbb{G}_n(f')\big| > \epsilon\right) \leq \eta.$$

We can show that condition 1 of Van der Vaart (2000, Theorem 18.14) holds for the stochastic process $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ by the Importance-Weighted Martingale Central Limit Theorem (Theorem 10). Specifically, by Cramer Wold device, it is sufficient to show that for any $c = [c_1, c_2, \ldots, c_K] \in \mathbb{R}^K$ that

$$\sum_{k=1}^{K} c_k \mathbb{G}_n(f_k) \xrightarrow{D} \mathcal{N}\left(0, c^\top \begin{bmatrix} Z_{1,1} & Z_{1,2} & \ldots & Z_{1,K} \\ Z_{2,1} & Z_{2,2} & \ldots & Z_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{K,1} & Z_{K,2} & \ldots & Z_{K,K} \end{bmatrix} c\right)$$

where $Z_{k,k'} = \mathbb{E}_{\pi^*_{2:T}}\left[\mathbb{G}_{\mathcal{F}}(f_k)\mathbb{G}_{\mathcal{F}}(f_{k'})\right]$. Note that

$$\sum_{k=1}^{K} c_k \mathbb{G}_n(f_k) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) - \mathbb{E}\left[\left(\hat{\pi}^{(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) \right] \right\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ W^{(i)}_{2:T}\left(\theta^*_{1:T-1}, \hat{\theta}^{(n)}_{1:T-1}\right)\left(\pi^{*,(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) \right.$$
$$\left. - \mathbb{E}\left[W^{(i)}_{2:T}\left(\theta^*_{1:T-1}, \hat{\theta}^{(n)}_{1:T-1}\right)\left(\pi^{*,(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) \right] \right\} \xrightarrow{D} \mathcal{N}\left(0, \Sigma_{\mathcal{F}}\right).$$

The above limit holds by the Importance-Weighted Martingale Central Limit Theorem (Theorem 10), for $\Sigma_{\mathcal{F}} \triangleq \mathbb{E}_{\pi^*_{2:T}}\left[\left\{ \left(\pi^{*,(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) \right\}^2 \right] - \mathbb{E}_{\pi^*_{2:T}}\left[\left(\pi^{*,(i)}_{2:T}\right)^{-1} \sum_{k=1}^{K} c_k f_k(\mathcal{H}^{(i)}_T) \right]^2$.

Note that $\Sigma_{\mathcal{F}} = c^\top \begin{bmatrix} Z_{1,1} & Z_{1,2} & \ldots & Z_{1,K} \\ Z_{2,1} & Z_{2,2} & \ldots & Z_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{K,1} & Z_{K,2} & \ldots & Z_{K,K} \end{bmatrix} c.$

The asymptotically tight condition above holds by the same argument used in the proof of Van der Vaart (2000, Theorem 19.5), but by replacing the use of maximal inequality Van der Vaart (2000, Lemma 19.34) in that proof with our maximal inequality from Lemma 14. ■

D.5. Lemma 16: Uniform Replacement of $\hat{\theta}^{(n)}$

Lemma 16 *We consider the setting in which the data is generated using the procedure described earlier in Section 1. Assuming* $\|\hat{\theta}^{(n)}_t(\,\cdot\,) - \theta^*_t(\,\cdot\,)\|_{\Theta_{1:t}} \xrightarrow{P} 0$, $\hat{\theta}^{(n)}_s \xrightarrow{P} \theta^*_s$ *for all* $s \in [1\colon t-1]$, *and under Conditions 1, 2, 3, 9, and 10 we have that for any fixed* $c_t \in \mathbb{R}^{d_t}$,

$$\Big\| \sqrt{n} c_t^\top \left[\hat{\Psi}^{(n)}_t\big(\,\cdot\,, \hat{\theta}^{(n)}_t(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \hat{\theta}^{(n)}_t(\,\cdot\,)\big) \right]$$
$$- \sqrt{n} c_t^\top \left[\hat{\Psi}^{(n)}_t\big(\,\cdot\,, \theta^*_t(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \theta^*_t(\,\cdot\,)\big) \right] \Big\|_{\Theta_{1:t-1}} \xrightarrow{P} 0. \quad (95)$$

Proof of Lemma 16 For this proof, we use an argument similar to Van der Vaart (2000, Lemma 19.24). We use $l^\infty(\mathcal{F})$ to refer to the collection of all bounded functions from $\mathcal{F}$ to $\mathbb{R}$.

Note that as discussed in the beginning of this section (Appendix D), by our choice of $\mathcal{F}_{t,c_t} \triangleq \big\{\big(\prod_{s=2}^{t-1}\pi_s(\,\cdot\,;\theta_{s-1})\big)c_t^\top\psi_t\big(\,\cdot\,;\theta_t\big) : \theta_s \in \Theta_s \text{ for all } s \in [1\colon t]\big\}$ (see Equation (65)), we have that

$$\Big\{\sqrt{n}c_t^\top\Big[\hat{\Psi}_t^{(n)}\big(\theta_{1:t}\big) - \Psi_t\big(\theta_{1:t}\big)\Big] : \theta_s \in \Theta_s \text{ for all } \in [1\colon t]\Big\}$$
$$= \Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1}f(\mathcal{H}_t^{(i)}) - \mathbb{E}\Big[\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1}f(\mathcal{H}_t^{(i)})\Big]\Big) : f \in \mathcal{F}_{t,c_t}\Big\}.$$

We let $f_{\theta_{1:t}}(\,\cdot\,) \triangleq \big(\prod_{s=2}^{t-1}\pi_s(\,\cdot\,;\theta_{s-1})\big)c_t^\top\psi_t\big(\,\cdot\,;\theta_t\big)$ and $\mathcal{F} = \{f_{\theta_{1:t}} : \theta_{1:t} \in \Theta_{1:t}\}$. We also let $\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}(f_{\theta_{1:t}}) \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1}f_{\theta_{1:t}}(\mathcal{H}_t^{(i)}) - \mathbb{E}\Big[\big(\hat{\pi}_{2:t}^{(i)}\big)^{-1}f_{\theta_{1:t}}(\mathcal{H}_t^{(i)})\Big]\Big)$ and $\hat{\mathbb{G}}_{\mathcal{F}}^{(n)} \triangleq \big\{\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}(f_{\theta_{1:t}}) : \theta_{1:t} \in \Theta_{1:t}\big\}$. We also let $\bar{\Theta}_t$ be the class of functions $\bar{\Theta}_t \triangleq \{\theta_t(\,\cdot\,) : \Theta_{1:t-1} \mapsto \Theta_t\}$.

By our Functional Asymptotic Normality under Finite Bracketing Integral result (Theorem 15), we have that the stochastic process $\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}$ converges in distribution to a mean-zero Gaussian process $\mathbb{G}_{\mathcal{F}}$ in $l^\infty(\mathcal{F})$ with covariance function

$$\mathbb{E}\big[\mathbb{G}_{\mathcal{F}}(f)\mathbb{G}_{\mathcal{F}}(g)\big] \triangleq \mathbb{E}_{\pi_{2:t}^*}\Big[\big(\pi_{2:t}^{*,(i)}\big)^{-2}f(\mathcal{H}_t^{(i)})g(\mathcal{H}_t^{(i)})\Big]$$
$$- \mathbb{E}_{\pi_{2:t}^*}\Big[\big(\pi_{2:t}^{*,(i)}\big)^{-1}f(\mathcal{H}_t^{(i)})\Big]\mathbb{E}_{\pi_{2:t}^*}\Big[\big(\pi_{2:t}^{*,(i)}\big)^{-1}g(\mathcal{H}_t^{(i)})\Big].$$

Condition 10 states that for any $\epsilon > 0$, there must exist a $\delta_{\epsilon,t} > 0$ such that for all $\theta_{1:t} \in \Theta_{1:t}$ with $\|\theta_{1:t} - \theta_{1:t}^*\| < \delta_{\epsilon,t}$, then $\rho_t(f_{\theta_{1:t}}, f_{\theta_{1:t}^*}) < \epsilon$. Recall that $\rho_t(f,f') \triangleq \mathbb{E}_{\pi_{2:t}^*}\big[\big\{f(\mathcal{H}_t^{(i)}) - f'(\mathcal{H}_t^{(i)})\big\}^2\big]$. Thus, for any function $\theta_t(\,\cdot\,) \in \bar{\Theta}_t$ with $\|\theta_t(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{\Theta_{1:t-1}} < \delta_{\epsilon,t}$ then we have that $\big\|\rho_t\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}, f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big)\big\|_{\Theta_{1:t-1}} < \epsilon$. Since $\|\hat{\theta}_t^{(n)}(\,\cdot\,) - \theta_t^*(\,\cdot\,)\|_{\Theta_{1:t}} \xrightarrow{P} 0$ by assumption, thus

$$\Big\|\rho_t\big(f_{(\,\cdot\,),\hat{\theta}_t^{(n)}(\,\cdot\,)}, f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big)\Big\|_{\Theta_{1:t-1}} \xrightarrow{P} 0$$

We will use the norm $\|\theta_t(\,\cdot\,) - \theta_t'(\,\cdot\,)\|_{\rho_t,\Theta_{1:t-1}} \triangleq \Big\|\rho_t\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}, f_{(\,\cdot\,),\theta_t'(\,\cdot\,)}\big)\Big\|_{\Theta_{1:t-1}}$ on $\theta_t(\,\cdot\,), \theta_t'(\,\cdot\,) \in \bar{\Theta}_t$. By the previous two results, Slutsky's Theorem implies that $\Big(\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}, \hat{\theta}_t^{(n)}(\,\cdot\,)\Big) \xrightarrow{D} (\mathbb{G}_{\mathcal{F}}, \theta_t^*(\,\cdot\,))$ in $\ell^\infty(\mathcal{F}) \times \bar{\Theta}_t$.

Consider the mapping that takes $\mathbb{G} \in \ell^\infty(\mathcal{F})$ and $\theta_t(\,\cdot\,) \in \bar{\Theta}_t$ and outputs $\mathbb{G}\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}\big) - \mathbb{G}\big(f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big) \in \ell^\infty(\Theta_{1:t-1})$. Let g be this mapping. We can write $g : l^\infty(\mathcal{F}) \times \bar{\Theta}_t \mapsto \ell^\infty(\Theta_{1:t-1})$ such that $g(z,h) \triangleq z(h) - z(h^*)$, where $z \in l^\infty(\mathcal{F})$ and $h, h^* \in \bar{\Theta}$.

- Note that $g(z,h) \in \ell^\infty(\Theta_{1:t-1})$ is continuous at a point $(z,h) \in \big(\ell^\infty(\mathcal{F}), \bar{\Theta}_t\big)$ if z is continuous in h at that point. By z being continuous in h at that point, we mean that for any $\epsilon > 0$, there exists some $\delta_\epsilon > 0$ such that $\|z(h) - z(h')\|_{\Theta_{1:t-1}} < \epsilon$ when ever

$\|h - h'\|_{\rho_t,\Theta_{1:t-1}} < \delta_\epsilon$. By $g(z,h)$ being continuous at a point $(z,h) \in \big(\ell^\infty(\mathcal{F}), \bar{\Theta}_t\big)$ we mean that for any $\epsilon > 0$, there exists some $\delta_\epsilon > 0$ such that $\|g(z,h) - g(z',h')\|_{\Theta_{1:t-1}} < \epsilon$ whenever $\sqrt{\|h - h'\|^2_{\rho_t,\Theta_{1:t-1}} + \|z - z'\|^2_{\Theta_{1:t-1}}} < \delta_\epsilon$. We provide a quick proof of this.

Let $\epsilon > 0$. Let $z' \in \ell^\infty(\mathcal{F})$ and $h \in \bar{\Theta}_t$. We assume z is continuous at h, so for some $\delta_{\epsilon/4} > 0$, $\|z(h) - z(h')\|_{\Theta_{1:t-1}} \leq \epsilon/4$ when ever $\|h - h'\|_{\rho_t,\Theta_{1:t-1}} < \delta_{\epsilon/4}$. We assume that $\sqrt{\|h - h'\|^2_{\rho_t,\Theta_{1:t-1}} + \|z - z'\|^2_{\Theta_{1:t-1}}} \leq \min(\delta_{\epsilon/4}, \epsilon/4) = \lambda_\epsilon$. Now note that

$$\begin{aligned}\big\|g(z,h) - g(z',h')\big\|_{\Theta_{1:t-1}} &\leq \big\|g(z,h) - g(z,h')\big\|_{\Theta_{1:t-1}} + \big\|g(z,h') - g(z',h')\big\|_{\Theta_{1:t-1}} \\ &\leq \big\|z(h) - z(h')\big\|_{\Theta_{1:t-1}} + \big\|z(h') - z'(h')\big\|_{\Theta_{1:t-1}} + \big\|z(h^*) - z'(h^*)\big\|_{\Theta_{1:t-1}}\end{aligned}$$

Since $\|h - h'\|_{\rho_t,\Theta_{1:t-1}} \leq \lambda_\epsilon \leq \delta_{\epsilon/4}$ by our continuity assumption, $\|z(h) - z(h')\|_{\Theta_{1:t-1}} < \epsilon/4$. Also note that $\sup_{h\in\bar{\Theta}} \big\|z(h) - z'(h)\big\|_{\Theta_{1:t-1}} \leq \|z - z'\|_{\Theta_{1:t-1}}$.

$$\leq \epsilon/4 + 2\|z - z'\|_{\Theta_{1:t-1}} \leq 3/4\epsilon < \epsilon.$$

The final inequality above holds because $\|z - z'\|_{\Theta_{1:t-1}} \leq \lambda_{\min} \leq \epsilon/4$. Thus we have shown our desired result.

Thus, to show that $g(z,h)$ is continuous at the point $(\mathbb{G}_\mathcal{F}, \theta_t(\,\cdot\,))$ it is sufficient to show that $\mathbb{G}_\mathcal{F}\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}\big) \in \ell^\infty(\Theta_{1:t-1})$ is continuous in $\theta_t(\,\cdot\,)$ at the point $\theta_t(\,\cdot\,) \in \bar{\Theta}_t$.

- We let $\mathbb{G}_\mathcal{F}[\omega]$ denote a sample path of $\mathbb{G}_\mathcal{F}$. By Lemma 18.15 of (Van der Vaart, 2000), almost all sample paths of $\mathbb{G}$ are continuous on $\mathcal{F}$, i.e., for almost all ω, for any $\epsilon > 0$, there exists some $\delta_\epsilon > 0$ such that $\big|\mathbb{G}_\mathcal{F}[\omega]\big(f\big) - \mathbb{G}_\mathcal{F}[\omega]\big(f'\big)\big| < \epsilon$ for any $f, f' \in \mathcal{F}$ with $sd(f - f') \triangleq \mathbb{E}\big[\mathbb{G}(f - f')\mathbb{G}(f - f')\big]^{1/2} \leq \delta_\epsilon$.

The previous result means that for almost all ω, for any $\epsilon > 0$, there exists some $\delta_\epsilon > 0$ such that $\big\|\mathbb{G}_\mathcal{F}[\omega]\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}\big) - \mathbb{G}_\mathcal{F}[\omega]\big(f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}\big)\big\|_{\Theta_{1:t-1}} < \epsilon$ for any $\theta_t(\,\cdot\,), \theta'_t(\,\cdot\,) \in \bar{\Theta}_t$ with $\big\|sd\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)} - f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}\,\big)\big\|_{\Theta_{1:t-1}} \leq \delta_\epsilon$.

Note that

$$\begin{aligned}&\Big\|sd\Big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}, f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}\Big)\Big\|_{\Theta_{1:t-1}} \\ &\qquad\leq \left\|\mathbb{E}_{\pi^*_{2:t}}\left[\big(\pi^{*,(i)}_{2:t}\big)^{-2}\left\{f_{(\,\cdot\,),\theta_t(\,\cdot\,)}(\mathcal{H}^{(i)}_t) - f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}(\mathcal{H}^{(i)}_t)\right\}^2\right]\right\|_{\Theta_{1:t-1}} \\ &\qquad\leq \pi^{-2t}_{\min}\Big\|\rho_t\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}, f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}\big)\Big\|_{\Theta_{1:t-1}} = \pi^{-2t}_{\min}\big\|\theta_t(\,\cdot\,) - \theta'_t(\,\cdot\,)\big\|_{\rho_t,\Theta_{1:t-1}}.\end{aligned}$$

Thus, for any $\epsilon > 0$, there exists some $\delta'_\epsilon > 0$ such that $\big\|\mathbb{G}_\mathcal{F}[\omega]\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}\big) - \mathbb{G}_\mathcal{F}[\omega]\big(f_{(\,\cdot\,),\theta'_t(\,\cdot\,)}\big)\big\|_{\Theta_{1:t-1}}$ $< \epsilon$ for any $\theta_t(\,\cdot\,), \theta'_t(\,\cdot\,) \in \bar{\Theta}_t$ with $\|\theta_t(\,\cdot\,) - \theta'_t(\,\cdot\,)\|_{\rho_t,\Theta_{1:t-1}} < \delta'_\epsilon$. So, almost all $\mathbb{G}_\mathcal{F}\big(f_{(\,\cdot\,),\theta_t(\,\cdot\,)}\big) \in \ell^\infty(\mathcal{F})$ are continuous in $\theta_t(\,\cdot\,)$ for all $\theta_t(\,\cdot\,) \in \bar{\Theta}_t$, where we use the sup norm on $\ell^\infty(\mathcal{F})$ and we use the norm $\|\theta_t(\,\cdot\,) - \theta'_t(\,\cdot\,)\|_{\rho_t,\Theta_{1:t-1}}$ on $\theta_t(\,\cdot\,), \theta'_t(\,\cdot\,) \in \bar{\Theta}_t$.

Thus, we can apply continuous mapping theorem to conclude that $g\left(\mathbb{G}_{\mathcal{F}}, \hat{\theta}_t^{(n)}(\,\cdot\,)\right)$ $\xrightarrow{D}$ $g\left(\mathbb{G}_{\mathcal{F}}, \theta_t^*(\,\cdot\,)\right)$, which means $\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\hat{\theta}_t^{(n)}(\,\cdot\,)}\big) - \hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big) \xrightarrow{D} 0$ in $l^\infty(\Theta_{1:t-1})$. Note that convergence in distribution to 0 implies convergence in probability to 0. Thus, $\big\|\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\hat{\theta}_t^{(n)}(\,\cdot\,)}\big) - \hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big)\big\|_{\Theta_{1:t-1}} \xrightarrow{P} 0$. This implies Equation (95) holds since

$$\left\|\hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\hat{\theta}_t^{(n)}(\,\cdot\,)}\big) - \hat{\mathbb{G}}_{\mathcal{F}}^{(n)}\big(f_{(\,\cdot\,),\theta_t^*(\,\cdot\,)}\big)\right\|_{\Theta_{1:t-1}}$$

$$= \Bigg\| \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \hat{\theta}_t^{(n)}(\,\cdot\,)\big)\right] - \sqrt{n} c_t^\top \left[\hat{\Psi}_t^{(n)}\big(\,\cdot\,, \theta_t^*(\,\cdot\,)\big) - \Psi_t\big(\,\cdot\,, \theta_t^*(\,\cdot\,)\big)\right] \Bigg\|_{\Theta_{1:t-1}}. \quad \blacksquare$$