A Spectral Approach to Off-Policy Evaluation for POMDPs

Yash Nair [1] **Nan Jiang** [2]

Abstract

We consider the off-policy evaluation problem in POMDPs. Prior work on this problem uses a causal identification strategy based on one-step observable proxies of the hidden state (Tennenholtz et al., 2020a). In this work, we relax the assumptions made in the prior work by using spectral methods. We further relax these assumptions by extending one-step proxies into the past. Finally, we derive an importance sampling algorithm which assumes rank, distinctness, and positivity conditions on certain probability matrices, and not on sufficiency conditions of observable trajectories with respect to the reward and hidden state structure required in the prior work.

1. Introduction

We consider the problem of estimating the value of an *evaluation policy* given access only to a batch of trajectories obtained from the *behavior policy*, known as the *off-policy evaluation* (OPE) problem. The OPE problem is well-motivated by its application to real-world scenarios in which the deployment of the evaluation policy is potentially too costly for its performance to be determined by directly intervening with it, but rather must be inferred from previously existing data under a different policy. As an example, in healthcare settings, it may be unsafe to directly test a new experimental treatment with potentially unknown side-effects; thus reasoning based on previous treatment strategies to infer a new treatment's efficacy is necessary.

We consider the regime of Partially Observable Markov Decision Processes (POMDPs), in which one has access only to *observable* trajectories under the behavior policy, which do not include underlying latent states (i.e. hidden confounders), and must infer the value of the evaluation policy given only these observables. In particular, just as in (Tennenholtz et al., 2020a), we consider the setting in which the behavior policy depends on the latent state, whereas the evaluation policy at any given time step depends only on the observed history up to that time step. This choice is motivated, for example, by the medical treatment scenario described in (Tennenholtz et al., 2020a): one must determine the effectiveness of the evaluation treatment, given access to the behavior treatment administered by a doctor who had access to—and thus, acted according to—unobserved confounders such as the patient's socioeconomic status and insurance coverage. This choice in the structure of behavior and evaluation policies captures the realistic possibility that historical data may depend on hidden confounders which we are unable to observe or act upon.

In this work, we first generalize the method of (Tennenholtz et al., 2020a). Motivated by spectral learning in Predictive State Representations (PSRs) and Hidden Markov Models (HMMs) (Boots et al., 2011; Kulesza et al., 2015; Hsu et al., 2012), we give an estimator whose rank assumptions scale with the size of the hidden state space. We then give further weaker assumptions by using *multi-step proxies* for hidden state, rather than the one-step proxies of (Tennenholtz et al., 2020a), by using the entire extended history preceding the confounder as one of the proxies, rather than just the previous time step. Along the way, we show that there is difficulty in using extended futures for the other proxy (which would further relax rank assumptions). Finally, by extending the eigendecomposition technique of (Kuroki & Pearl, 2014), we provide an Importance Sampling (IS) algorithm which requires only rank conditions on certain probability matrices and not the sufficiency assumptions of (Tennenholtz et al., 2020a).

2. Related Work

OPE in POMDPs Most closely related to our work is that of (Tennenholtz et al., 2020a), who consider the finite-horizon OPE problem in tabular POMDPs by extending the work of (Miao et al., 2018) and writing the probability of seeing reward r_t, in terms of matrices of observable probabilities, whose dimensions scale with the size of the observation space, $\mathcal{Z}$, and which are required to be invertible. They further develop a Decoupled POMDP model which factors the state space, $\mathcal{U}$, into both observed and unobserved variables, allowing for a modified OPE algorithm

[1]Harvard University [2]University of Illinois at Urbana-Champaign. Correspondence to: Yash Nair <yash-nair@college.harvard.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

whose observable matrices scale with $|\mathcal{U}|$, which is often much smaller than $|\mathcal{Z}|$. Our work mitigates difficulty in the general POMDP setting by requiring that the observable matrices have rank $|\mathcal{U}|$, rather than $|\mathcal{Z}|$, and further relaxes assumptions by extending histories.

Causal effect identification Our work is closely related to the use of negative controls to minimize confounding bias in the causal inference literature. Of particular interest is (Miao et al., 2018) and (Kuroki & Pearl, 2014), who consider the static problem of identifying the causal effect of one variable on another, assuming multiple observable proxies of the latent confounder. (Kuroki & Pearl, 2014) first identify the confounder's error mechanism via eigenvalue analysis and then use the matrix adjustment method (Greenland & Lash, 2012) to determine the causal effect, while (Miao et al., 2018) directly determine the causal effect without identifying any of the aspects of the confounding model, allowing for weaker assumptions. Our work generalizes the latter's results and extends to the setting of POMDPs, similar to (Tennenholtz et al., 2020a). We extend the former's analysis to give an IS algorithm which does not rely on sufficiency assumptions.

3. Preliminaries

POMDPs A POMDP is a 7-tuple $\langle \mathcal{U}, \mathcal{Z}, \mathcal{A}, O, T, R, H \rangle$, where $\mathcal{U}$ and $\mathcal{Z}$ denote the hidden state and observation spaces, respectively, $\mathcal{A}$ denotes the set of actions, O gives observation emission probabilities, T governs the hidden state transition dynamics, R is a (potentially non-deterministic) function of the hidden state and action giving the reward, and H denotes the horizon. In this work, we assume that the sets $\mathcal{U}$, $\mathcal{Z}$, and $\mathcal{A}$ are all finite.

Using the notation of (Tennenholtz et al., 2020a), a trajectory τ of length t denotes a sequence $(u_0, z_0, a_0, \ldots, u_t, z_t, a_t)$ while τ^o denotes the observable trajectory $(z_0, a_0, \ldots, z_t, a_t)$. We use $\mathcal{T}_t$ to denote the set of trajectories of length t, and, correspondingly, $\mathcal{T}_t^o$ to denote the set of observable trajectories of length t. Additionally, an observable history of length t is given by a sequence of the form $(z_0, a_0, \ldots, a_{t-1}, z_t)$, and is denoted by h_t^o. Finally, we will assume the existence of an observation z_{-1} preceding the initial time step which is conditionally independent of z_0 and a_0 given u_0.

Policies Let π_b and π_e denote the behavior and evaluation policies, respectively. As mentioned in the introduction, we consider the setting in which the behavior policy at time step t, $\pi_b^{(t)}$, is a stochastic function of the hidden state, u, whereas the evaluation policy at time t, $\pi_e^{(t)}$, is a stochastic function of only the observable history, h_t^o. Additionally, we consider the finite-horizon undiscounted setting so that for any policy π (either depending on observable histories or hidden state), its value $v_H(\pi)$ is given by $\mathbb{E}_\pi\left[\sum_{t=0}^{H} r_t\right]$, the expected sum of rewards seen by following π. Finally, let P^b and P^e denote measures over trajectories induced by the behavior and evaluation policies, respectively.

In addition to the above, we also use the double vertical bar notation of (Boots et al., 2011) to indicate intervening. For example, $P[z_0, \ldots, z_t || a_0, \ldots, a_{t-1}]$ denotes the probability of seeing the observation sequence $(z_0, \ldots, z_t)$, given that we intervened with actions $a_0, \ldots, a_{t-1}$.

Problem Formally, the OPE problem asks, given a batch of observable trajectories collected under the behavior policy, to estimate the value of the evaluation policy. In other words, we must estimate $v_H(\pi_e)$ given only observable trajectories from P^b.

For convenience, we use the vector notations of (Tennenholtz et al., 2020a). For example, let the random variables $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be supported on the sets $X = \{x_1, \ldots, x_{n_1}\}, Y = \{y_1, \ldots, y_{n_2}\}, Z = \{z_1, \ldots, z_{n_3}\}$, respectively. Then $P(y|X), P(X), P(Y|x, Z)$ denote a row vector, column vector, and matrix, respectively with $(P(y|X))_i = \Pr[y|x_i], (P(X))_i = \Pr[x_i], ((P(Y|x, Z))_{i,j} = P(y_i|x, z_j)$. In general, a set (always denoted by a capital letter) before (after) the conditioning bar indicates vectorization across rows (columns).

3.1. One-Step Proxies

We now briefly review the POMDP OPE algorithm of (Tennenholtz et al., 2020a), who approach the OPE problem by directly estimating $P^e(r_t)$ which can then be used to estimate $v_H(\pi_e)$. They make the following assumptions on observable probability matrices.

Assumption 1. *$P^b(Z_i|a_i, Z_{i-1})$ is invertible $\forall i \leq H, a_i \in \mathcal{A}$.*

Defining $\Pi_t^e(\tau^o) = \prod_{i=0}^{t} \pi_e^{(i)}(a_i|h_i^o)$, they give the following estimator:

Theorem 1. *Under Assumption 1, $P^e(r_t)$ can be written as*

$$\sum_{\tau^o \in \mathcal{T}_t^o} \Pi_e(\tau^o) P^b(r_t, z_t | a_t, Z_{t-1}) \cdot \prod_{i=t-1}^{0} \left(P^b(Z_{i+1}|a_{i+1}, Z_i)^{-1} P^b(Z_{i+1}, z_i | a_i, Z_{i-1}) \right) \cdot P^b(Z_0|a_0, Z_{-1})^{-1} P^b(Z_0). \quad (1)$$

The proof is made possible by first noting that

$$P^e(r_t) = \sum_{\tau^o \in \mathcal{T}_t^o} \Pi_t^e(\tau^o) P^b(r_t, z_t, z_{t-1}, \ldots, z_0 || a_0, \ldots, a_t), \quad (2)$$

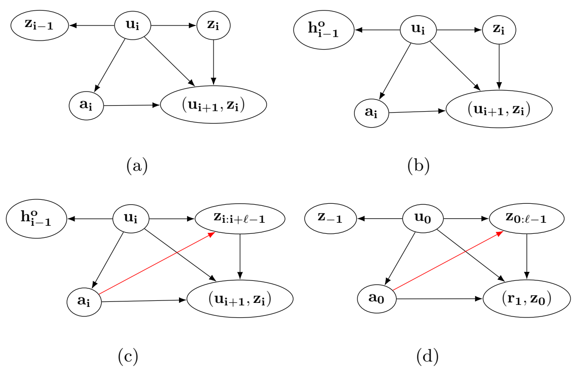

Figure 1. Causal diagrams for proxies of the confounder $\mathbf{u_i}$: one-step proxies (a), multi-step histories (b), multi-step histories, multi-step futures (c), and multi-step futures with reward-observation outcome (d).

and then, at each time step i, statically viewing z_{i-1} as an observable proxy emitted by the latent state u_i, giving the causal diagram in Figure 1(a) which implies that the proxies z_i and z_{i-1} are conditionally independent given u_i, allowing for the application of the identification scheme of (Miao et al., 2018) to identify $P^b(r_t, z_t, z_{t-1}, \dots, z_0 || a_0, \dots, a_t)$.

4. Relaxing Rank Assumptions

As $P^b(Z_i|a_i, Z_{i-1}) = P^b(Z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1})$, Assumption 1 implicitly requires $|\mathcal{U}| \geq |\mathcal{Z}|$ (since for matrices, A, B, $\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))$), which is unrealistic in most real-world settings, where the hidden state space is often much smaller than the observation space. Inspired by spectral learning of PSRs and POMDPs (Boots et al., 2011; Kulesza et al., 2015; Hsu et al., 2012), we show that we can allow for the more reasonable assumption that $|\mathcal{U}| \leq |\mathcal{Z}|$ by requiring the matrices $P^b(Z_i|a_i, Z_{i-1})$ to be of rank $|\mathcal{U}|$. As such, we henceforth assume that $|\mathcal{U}| \leq |\mathcal{Z}|$. We then relax assumptions by using the entire observable history as a proxy, rather than just the observation from the previous time step.

4.1. Spectral Relaxation of One-Step Proxies

We first relax rank assumptions in the case of one-step proxies by performing singular value decompositions (SVDs) on the matrices $P^b(Z_i|a_i, Z_{i-1})$ and using the left singular vectors to derive spectral analogs to the causal identification lemmas used in Theorem 1. We make the following assumption:

Assumption 2. $\text{rank}(P^b(Z_i|a_i, Z_{i-1})) \geq |\mathcal{U}|, \forall i \leq H, a_i \in \mathcal{A}$.

Under Assumption 2, we obtain the following estimator (see Appendix for proof):

Theorem 2. *Under Assumption 2, let M_{i,a_i} be the left singular vectors of $P^b(Z_i|a_i, Z_{i-1})$ and $M'_{i,a_i} := (M_{a_i}P(Z_i|a_i, Z_{i-1}))^+$. Then $P^e(r_t)$ is equal to*

$$\sum_{\tau^o \in \mathcal{T}^o_t} \Pi_e(\tau^o) P^b(r_t, z_t|a_t, Z_{t-1}) M'_{t,a_t} \cdot \prod_{i=t-1}^{0} \left(M_{i+1,a_{i+1}} P^b(Z_{i+1}, z_i|a_i, Z_{i-1}) M'_{i,a_i}\right) M_{0,a_0} P^b(Z_0) \tag{3}$$

4.2. Multi-Step Histories

We now show that, by extending histories, even weaker rank assumptions can be made. Specifically, letting $\mathcal{H}^o_t$ denote the set of observable histories of length t, the matrices, $P^b(Z_i|a_i, Z_{i-1})$ and $P^b(Z_{i+1}, z_i|a_i, Z_{i-1})$, in Theorem 2 can be replaced with $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$ and $P^b(Z_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})$, respectively, with M_{i,a_i} and M'_{i,a_i} defined correspondingly in terms of SVDs of $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$. We make the following rank assumption:

Assumption 3. $\text{rank}(P^b(Z_i|a_i, \mathcal{H}^o_{i-1})) \geq |\mathcal{U}|, \forall i \leq H, a_i \in \mathcal{A}$.

Similar to Theorem 2, we obtain the following:

Theorem 3. *Under Assumption 3, define M_{i,a_i} and M'_{i,a_i} in terms of $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$. Then the probability $P^e(r_t)$ is equal to*

$$\sum_{\tau^o \in \mathcal{T}^o_t} \Pi_e(\tau^o) P^b(r_t, z_t|a_t, \mathcal{H}^o_{t-1}) M'_{t,a_t} \cdot \prod_{i=t-1}^{0} \left(M_{i+1,a_{i+1}} P^b(Z_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1}) M'_{i,a_i}\right) M_{0,a_0} P^b(Z_0) \tag{4}$$

We give a proof of Theorem 3 in the Appendix, but, at a high level, Figures 1(a) and (b) show the static causal structure is unchanged when extending the length of observable histories, allowing for essentially the same analysis as in the one-step case.

Importantly, note that Assumption 3 is weaker than Assumption 2 as $P^b(Z_i|a_i, Z_{i-1}) = P^b(Z_i|a_i, \mathcal{H}^o_{i-1})P^b(\mathcal{H}^o_{i-1}|a_i, Z_{i-1})$.

5. Multi-Step Futures

We would like to extend futures and replace the matrices $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$ and $P^b(Z_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})$ with probability matrices of length ℓ futures of the form $P^b(Z_{i:i+\ell}|a_i, \mathcal{H}^o_{i-1})$ and $P^b(Z_{i+1:i+1+\ell}, z_i|a_i, \mathcal{H}^o_{i-1})$ (where $Z_{i:j} := \prod_{k=i}^{j} Z_k$ and $z_{i:j} := (z_i, \dots, z_j)$ for $i \leq j$; when $j < i$, these are the empty set and sequence, respectively), so as to further relax rank assumptions by only assuming that rank$(P^b(Z_{i:i+\ell}|a_i, \mathcal{H}^o_{i-1})) \geq |\mathcal{U}|$, which is implied by Assumption 3. This, however, as we now show, is not possible via identification of $P^b(r_t, z_t, z_{t-1}, \ldots, z_0||a_0, \ldots, a_t)$.

The primary difficulty, as indicated by the red edge in Figure 1(c), is that, when extending futures, the treatment, a_i, *does* have a causal effect on the observables $z_{i:i+\ell-1}$, yielding a causal structure in which the causal effect is nonidentifiable (Pearl, 2009). Hence, the above identification strategy is destined to fail when using multi-step futures as proxies of the confounder. In fact, it is, in general, impossible to write the probability $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$ (on whose identifiability our approach crucially depends via (2)) in terms of probabilities involving multi-step futures, $z_{i:i+\ell-1}$ (without, of course, simply marginalizing out $z_{i+1:i+\ell-1}$) as evidenced, for example, by Figure 1(d), which also has the same causal structure, thus implying that even the probability $P(r_1, z_0||a_0)$ is non-identifiable via multi-step future proxies.

6. Importance Sampling

We finally derive an IS procedure for OPE in POMDPs. While the assumptions for our IS procedure are not directly comparable with those made in previous sections, they involve only rank, distinctness, and positivity conditions on certain probabilities, and not the sufficiency assumptions made in (Tennenholtz et al., 2020a) wherein τ^o is assumed to be a sufficient statistic for both reward and state. Rather, we extend the eigendecomposition technique of (Kuroki & Pearl, 2014), which is used to calculate matrices of probabilities involving *unobserved* confounders.

In order to satisfy certain rank assumptions of our analysis we require:

Assumption 4. $|\mathcal{U}| \leq |\mathcal{A}|$.

Assumption 5. $|\mathcal{R}| \geq |\mathcal{U}|$, *where $\mathcal{R}$ denotes the support of possible rewards*

Under rank $|\mathcal{U}|$ conditions on certain matrices of observable probabilities, we are then able to write vectors of the form $\pi_b(a_i|U_i)$ via eigendecompositions of matrices of observable probabilities under the behavior measure, where the entries are subject to an unknown, but consistent, ordering (i.e., the kth entry of $\pi_b(a_i|U_i)$ is $\pi_b(a_i|u_i^{(k)})$, but with the identity of $u_i^{(k)}$ unknown). We, additionally, make a weak assumption on the reward distribution so that, the causal structure of rewards and observations are identical:

Assumption 6. *The reward function R is a (possibly stochastic) function of only latent state, and does not depend on action.*

Assumption 6 allows for the reward probability distribution $P^b(r_i|U_i)$ to also be written in terms of observable probabilities in the same way as described above and under similar rank $|\mathcal{U}|$ assumptions. Additionally, in order to ensure uniqueness in the eigendecompositions, we further require certain probabilities to be distinct. As these conditions are in terms of probability matrices which lack succinct representation, we relegate the assumptions and proof to the Appendix but summarize here:

Theorem 4. *Under Assumptions 4, 5, and 6, as well as rank $|\mathcal{U}|$ and distinctness conditions on certain probability matrices, the vectors $\pi_b^{(i)}(a_i|U_i)$, and $P^b(r_i|U_i)$ can be written in terms of observable probabilities under the behavior measure with respect to an unknown, but consistent, ordering on U_i for all $i \leq H$.*

Defining $\Pi_b(u_{0:t}) = \prod_{i=0}^t \pi_b^{(i)}(a_i|u_i)$ and letting v denote value, we then obtain an IS procedure under the two following additional exponential rank and positivity assumptions, as well as additional probability positivity assumptions (see Appendix):

Assumption 7. rank$(P^b(U_{0:H}|\mathcal{T}^o_H)) = |\mathcal{U}|^{H+1}$

Assumption 8. $P^b(v|\tau^o) > 0, \forall v, \tau^o$ *and* $\pi_b^{(i)}(a_i|u_i) > 0, \forall i \leq H, a_i \in \mathcal{A}, u_i \in \mathcal{U}$.

Theorem 5. *Under Assumptions 4, 5, 6, 7, 8, as well as additional rank, distinctness, and positivity conditions on certain observable probability matrices, we have that*

$$v_H(\pi_e) = \mathbb{E}_{\tau^o}\left[\mathbb{E}_v[vW_{e,b}(v, \tau^o)|\tau^o]\right]$$

with $W_{e,b}(v, \tau^o) := \frac{P^b(\tau^o)}{P^b(v|\tau^o)}\Pi_e(\tau^o)\Gamma_b(v, \tau^o)$ *and*

$$\Gamma_b(v, \tau^o) = \sum_{r_{0:H}:\sum r_i = v} \sum_{u_{0:H}} \frac{\prod_{i=0}^H P^b(r_i|u_i) P^b(u_{0:H}|\tau^o)}{\Pi_b(u_{0:H})},$$

where $W_{e,b}(v, \tau^o)$ is identifiable in terms of observable probabilities.

7. Conclusion

In this work we consider OPE in POMDPs under the framework posed in (Tennenholtz et al., 2020a). We extend their work, relaxing assumptions on observable probability matrices to only have rank $|\mathcal{U}|$ rather than $|\mathcal{Z}|$, simultaneously assuming that $|\mathcal{U}| \leq |\mathcal{Z}|$. By additionally extending one of the proxy variables into the past, we further relax assumptions by assuming the matrices $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$ have rank $|\mathcal{U}|$, allowing for our estimator to apply to a broader class of POMDPs. We show that futures, however, cannot be extended using this strategy, and additionally cannot, in general, be used to obtain probabilities of the form $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$. Finally, we give an IS algorithm for OPE in POMDPs which only depends on rank, distinctness, and positivity conditions on certain probability matrices and not on sufficiency assumptions.

References

Bennett, A., Kallus, N., Li, L., and Mousavi, A. Off-policy evaluation in infinite-horizon reinforcement learning with latent confounders, 2020.

Boots, B., Siddiqi, S. M., and Gordon, G. J. Closing the learning-planning loop with predictive state representations. *The International Journal of Robotics Research*, 30(7):954–966, 2011. doi: 10.1177/0278364911404092. URL https://doi.org/10.1177/0278364911404092.

Greenland, S. and Lash, T. Bias Analysis. *Modern Epedimology*, 3:423–437, 2012.

Hsu, D., Kakade, S. M., and Zhang, T. A spectral algorithm for learning hidden markov models. *Journal of Computer and System Sciences*, 78 (5):1460 – 1480, 2012. ISSN 0022-0000. doi: https://doi.org/10.1016/j.jcss.2011.12.025. URL http://www.sciencedirect.com/science/article/pii/S0022000012000244. JCSS Special Issue: Cloud Computing 2011.

Jiang, N. and Li, L. Doubly robust off-policy value evaluation for reinforcement learning. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 652–661, New York, New York, USA, 20–22 Jun 2016. PMLR. URL http://proceedings.mlr.press/v48/jiang16.html.

Kulesza, A., Jiang, N., and Singh, S. Spectral learning of predictive state representations with insufficient statistics. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pp. 2715–2721. AAAI Press, 2015. ISBN 0262511290.

Kuroki, M. and Pearl, J. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2): 423–437, 03 2014. ISSN 0006-3444. doi: 10.1093/biomet/ast066. URL https://doi.org/10.1093/biomet/ast066.

Mandel, T., Liu, Y.-E., Brunskill, E., and Popović, Z. Offline evaluation of online reinforcement learning algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

Oberst, M. and Sontag, D. Counterfactual off-policy evaluation with Gumbel-max structural causal models. In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 4881–4890, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL http://proceedings.mlr.press/v97/oberst19a.html.

Pearl, J. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.

Precup, D., Sutton, R. S., and Singh, S. P. Eligibility traces for off-policy policy evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, ICML '00, pp. 759–766, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc. ISBN 1558607072.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

Tennenholtz, G., Shalit, U., and Mannor, S. Off-policy evaluation in partially observable environments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10276–10283, Apr. 2020a. doi: 10.1609/aaai.v34i06.6590. URL https://ojs.aaai.org/index.php/AAAI/article/view/6590.

Tennenholtz, G., Shalit, U., and Mannor, S. Off-policy evaluation in partially observable environments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10276–10283, Apr. 2020b. doi: 10.1609/aaai.v34i06.6590. URL https://ojs.aaai.org/index.php/AAAI/article/view/6590.

Thomas, P. and Brunskill, E. Data-efficient off-policy policy evaluation for reinforcement learning. In Balcan, M. F. and Weinberger, K. Q. (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2139–2148, New York, New York, USA, 20–22 Jun 2016. PMLR. URL http://proceedings.mlr.press/v48/thomasa16.html.

Zhang, J. and Bareinboim, E. Markov decision processes with unobserved confounders : A causal approach. 2016.

A. OPE via Observable Proxies

We extend the identification scheme of (Tennenholtz et al., 2020a) and (Miao et al., 2018) to make weaker rank assumptions. The proofs in this section use the same general strategy as in (Tennenholtz et al., 2020a), but use spectral methods to relax their assumptions. We first give a sketch of the proof of 1 in (Tennenholtz et al., 2020a) before proceeding to our rank relaxations.

A.1. (Tennenholtz et al., 2020a) One-Step Proxies

Proof Sketch. Note that

$$P^e(r_t) = \sum_{\tau^o \in \mathcal{T}_t^o} \Pi_t^e(\tau^o) P^b(r_t, z_t, z_{t-1}, \dots, z_0 || a_0, \dots, a_t). \tag{5}$$

We will identify $P^b(r_t, z_t, z_{t-1}, \dots, z_0 || a_0, \dots, a_t)$. Essentially, the analysis rests on the ability to non-parametrically identify the causal effect of the control a_i on the outcome (u_{i+1}, z_i) given confounded proxy variables z_{i-1} and z_i at each time step using the procedure of (Miao et al., 2018).

More specifically, this is done by first noting that

$$P(r_t, z_t, z_{t-1}, \dots, z_0 || a_0, \dots, a_t) = P^b(r_t, z_t | a_t, U_t) \left(\prod_{i=t-1}^{0} P^b(U_{i+1}, z_i | a_i, U_i) \right) P^b(U_0), \tag{6}$$

and extending off of the static causal identification setting, to show that

$$\begin{aligned} &P^b(U_{i+1}, z_i | a_i, U_i) P^b(U_i, z_{i-1} | a_{i-1}, U_{i-1}) \\ &= P^b(U_{i+1}, z_i | a_i, Z_{i-1}) P^b(Z_i | a_i, Z_{i-1})^{-1} P^b(Z_i, z_{i-1} | a_{i-1}, Z_{i-2}) P^b(Z_{i-1} | a_{i-1}, Z_{i-2})^{-1} P^b(Z_{i-1} | a_{i-1}, U_{i-1}) \end{aligned} \tag{7}$$

and

$$\begin{aligned} &P^b(r_t, z_t | a_t, U_t) P^b(U_t, z_{t-1} | a_{t-1}, U_{t-1}) \\ &= P^b(r_t, z_t | a_t, Z_{t-1}) P^b(Z_t | a_t, Z_{t-1})^{-1} P^b(Z_t, z_{t-1} | a_{t-1}, Z_{t-2}) P^b(Z_{t-1} | a_{t-1}, Z_{t-2})^{-1} P^b(Z_{t-1} | a_{t-1}, U_{t-1}) \end{aligned} \tag{8}$$

while also observing that

$$P^b(Z_i | a_i, U_i) P^b(U_i, z_{i-1} | a_{i-1}, Z_{i-2}) = P^b(Z_i, z_{i-1} | a_{i-1}, Z_{i-2}). \tag{9}$$

From these 4 core identities, it is then not hard to inductively derive the result. □

A.2. Spectral One-Step Proxies

The proof of Theorem 2 relies on deriving the following analogs to equations (7) – (8) and using equations (6) and (9) to similarly proceed by induction.

Lemma 1. *Given Assumption 2, define M_{i,a_i} and M'_{i,a_i} as described in Section . Then*

$$\begin{aligned} &P^b(U_{i+1}, z_i | a_i, U_i) P^b(U_i, z_{i-1} | a_{i-1}, U_{i-1}) \\ &= P^b(U_{i+1}, z_i | a_i, Z_{i-1}) M'_{i,a_i} M_{i,a_i} P^b(Z_i, z_{i-1} | a_{i-1}, Z_{i-2}) \\ &\quad \cdot M'_{i-1,a_{i-1}} M_{i-1,a_{i-1}} P^b(Z_{i-1} | a_{i-1}, U_{i-1}) \end{aligned} \tag{10}$$

and

$$\begin{aligned} &P^b(r_t, z_t | a_t, U_t) P^b(U_t, z_{t-1} | a_{t-1}, U_{t-1}) \\ &= P^b(r_t, z_t | a_t, Z_{t-1}) M'_{t,a_t} M_{t,a_t} P^b(Z_t, z_{t-1} | a_{t-1}, Z_{t-2}) \\ &\quad \cdot M'_{t-1,a_{t-1}} M_{t-1,a_{t-1}} P^b(Z_{t-1} | a_{t-1}, U_{t-1}) \end{aligned} \tag{11}$$

A.2.1. Proof of Lemma 1

Proof. First, notice that

$$P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1}) = P^b(U_{i+1}, z_i|a_i, Z_{i-1})$$

$$\implies P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1})M'_{i,a_i} = P^b(U_{i+1}, z_i|a_i, Z_{i-1})M'_{i,a_i}. \tag{12}$$

Additionally,

$$P^b(Z_i|a_i, Z_{i-1}) = P^b(Z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1}) \tag{13}$$

$$\implies P^b(Z_i|a_i, Z_{i-1})M'_{i,a_i} = P^b(Z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1})M'_{i,a_i}. \tag{14}$$

From the definition of M'_{i,a_i}, we have that

$$\begin{aligned} &P^b(Z_i|a_i, Z_{i-1})M'_{i,a_i} \\ &= P^b(Z_i|a_i, Z_{i-1})(M_{i,a_i}P^b(Z_i|a_i, Z_{i-1}))^+ = P^b(Z_i|a_i, Z_{i-1})V_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+ \\ &= U_{i,a_i}\Sigma_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+, \end{aligned}$$

which has rank $|\mathcal{U}|$, thus implying, by equation (14), that rank $\left(P^b(U_i|a_i, Z_{i-1})M'_{i,a_i}\right) = |U|$, and so equation (12) implies that

$$P^b(U_{i+1}, z_i|a_i, U_i) = P^b(U_{i+1}, z_i|a_i, Z_{i-1})M'_{i,a_i}\left(P^b(U_i|a_i, Z_{i-1})M'_{i,a_i}\right)^{-1}. \tag{15}$$

Now, notice that equation (13) implies that

$$M_{i,a_i}P^b(Z_i|a_i, U_i)P^b(U_i|a_i, Z_{i-1}) = M_{i,a_i}P^b(Z_i|a_i, Z_{i-1}). \tag{16}$$

Similarly notice that rank $\left(M_{i,a_i}P^b(Z_i|a_i, Z_{i-1})\right) = |U|$ as, from the definition of M_{i,a_i}, we have $M_{i,a_i}P^b(Z_i|a_i, Z_{i-1}) = I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i}V^T_{i,a_i}$. This, in turn, implies that rank $\left(M_{i,a_i}P^b(Z_i|a_i, U_i)\right) = |U|$ by equation (16), and so equation (16) implies that

$$P^b(U_i|a_i, Z_{i-1}) = \left(M_{i,a_i}P^b(Z_i|a_i, U_i)\right)^{-1}M_{i,a_i}P^b(Z_i|a_i, Z_{i-1}),$$

implying that

$$P^b(U_i|a_i, Z_{i-1})M'_{i,a_i} = \left(M_{i,a_i}P^b(Z_i|a_i, U_i)\right)^{-1}M_{i,a_i}P^b(Z_i|a_i, Z_{i-1})M'_{i,a_i}. \tag{17}$$

Now, we claim that $M_{i,a_i}P^b(Z_i|a_i, Z_{i-1})M'_{i,a_i} = I_{|\mathcal{U}|\times|\mathcal{U}|}$ since

$$\begin{aligned} M_{i,a_i}P^b(Z_i|a_i, Z_{i-1})M'_{i,a_i} &= I_{|\mathcal{U}|\times|\mathcal{Z}|}U^T_{i,a_i}U_{i,a_i}\Sigma_{i,a_i}V^T_{i,a_i}\left(I_{|\mathcal{U}|\times|\mathcal{Z}|}U^T_{i,a_i}U_{i,a_i}\Sigma_{i,a_i}V^T_{i,a_i}\right)^+ \\ &= I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+ \end{aligned}$$

and thus equation (17) becomes

$$P^b(U_i|a_i, Z_{i-1})M'_{i,a_i} = \left(M_{i,a_i}P^b(Z_i|a_i, U_i)\right)^{-1}. \tag{18}$$

Finally, substituting (18) into (15) gives

$$\begin{aligned} &P^b(U_{i+1}, z_i|a_i, U_i) \\ &= P^b(U_{i+1}, z_i|a_i, Z_{i-1})M'_{i,a_i}M_{i,a_i}P^b(Z_i|a_i, U_i). \end{aligned}$$

This, in combination with the fact that

$$P^b(Z_i|a_i, U_i)P^b(U_i, z_{i-1}|a_{i-1}, Z_{i-2}) = P^b(Z_i, z_{i-1}|a_{i-1}, Z_{i-2}) \tag{19}$$

immediately implies that

$$\begin{aligned} P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i, z_{i-1}|a_i, U_i) &= P(U_{i+1}, z_i|a_i, Z_{i-1})M'_{i,a_i} \\ &\cdot M_{i,a_i}P^b(Z_i, z_{i-1}|a_{i-1}, Z_{i-2})M'_{a_{i-1}}M_{a_{i-1}}P^b(Z_{i-1}|a_{i-1}U_{i-1}), \end{aligned}$$

as desired. Replacing U_{i+1} with r_i gives the second part of the Lemma. □

The proof of Theorem 2 then follows first from writing $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$ as

$$P^b(r_t, z_t|a_t, U_t)\left(\prod_{i=t-1}^{0} P^b(U_{i+1}, z_i|a_i, U_i)\right) P^b(U_0),$$

using Lemma 1 and equation (19) above to inductively show that

$$\begin{aligned} &P^b(r_t, z_t|a_t, U_t)\left(\prod_{i=t-1}^{0} P^b(U_{i+1}, z_i|a_i, U_i)\right) P^b(U_0) \\ &\quad = P^b(r_t, z_t|a_t, Z_{t-1})M'_{t,a_t} \prod_{i=t-1}^{0} \left(M_{i+1,a_{i+1}} P^b(Z_{i+1}, z_i|a_i, Z_{i-1})M'_{i,a_i}\right) M_{0,a_0} P^b(Z_0), \end{aligned} \tag{20}$$

and then using the fact that

$$P^e(r_t) = \sum_{\tau^o \in \mathcal{T}_t^o} \Pi_e(\tau^o) P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$$

to obtain the final result.

A.3. Multi-Step Histories

A.3.1. Proof of Theorem 3

The proof of Theorem 3 is essentially the same as that of Theorem 2, except that we replace the one-step history Z_{i-1} with $\mathcal{H}^o_{i-1}$. For completeness, we include it below, first proving a Lemma analogous to Lemma 1:

Lemma 2. *Given Assumption 3, let $U_{i,a_i}\Sigma_{i,a_i}V_{i,a_i}^T$ be an SVD of $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$ with the diagonal entries of Σ_{i,a_i} in descending order and define $M_{i,a_i} := I_{|\mathcal{U}|\times|\mathcal{Z}|}U_{i,a_i}^T$ and $M'_{i,a_i} := (M_{a_i}\mathcal{H}^o_{i-1})^+$. Then*

$$\begin{aligned} &P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i, z_{i-1}|a_{i-1}, U_{i-1}) \\ &\quad = P^b(U_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i}M_{i,a_i}P^b(Z_i, z_{i-1}|a_{i-1}, \mathcal{H}^o_{i-2}) \\ &\qquad \cdot M'_{i-1,a_{i-1}}M_{i-1,a_{i-1}}P^b(Z_{i-1}|a_{i-1}, U_{i-1}) \end{aligned}$$

and

$$\begin{aligned} &P^b(r_t, z_t|a_t, U_t)P^b(U_t, z_{t-1}|a_{t-1}, U_{t-1}) \\ &\quad = P^b(r_t, z_t|a_t, \mathcal{H}^o_{t-1})M'_{t,a_t}M_{t,a_t}P^b(Z_t, z_{t-1}|a_{t-1}, \mathcal{H}^o_{t-2}) \\ &\qquad \cdot M'_{t-1,a_{t-1}}M_{t-1,a_{t-1}}P^b(Z_{t-1}|a_{t-1}, U_{t-1}) \end{aligned}$$

Proof. First, notice that

$$P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i|a_i, \mathcal{H}^o_{i-1}) = P^b(U_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})$$

$$\implies P^b(U_{i+1}, z_i|a_i, U_i)P^b(U_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i} = P^b(U_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i}. \tag{21}$$

Additionally,

$$P^b(Z_i|a_i, \mathcal{H}^o_{i-1}) = P^b(Z_i|a_i, U_i)P^b(U_i|a_i, \mathcal{H}^o_{i-1}) \tag{22}$$

$$\implies P^b(Z_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i} = P^b(Z_i|a_i, U_i)P^b(U_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i}. \tag{23}$$

From the definition of M'_{i,a_i}, we have that

$$\begin{aligned} &P^b(Z_i|a_i, \mathcal{H}^o_{i-1})M'_{i,a_i} \\ &= P^b(Z_i|a_i, \mathcal{H}^o_{i-1})(M_{i,a_i}P^b(Z_i|a_i, \mathcal{H}^o_{i-1}))^+ = P^b(Z_i|a_i, \mathcal{H}^o_{i-1})V_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+ \\ &= U_{i,a_i}\Sigma_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+ = U_{i,a_i}I_{|\mathcal{Z}\times|\mathcal{U}|} = U_{i,a_i}I_{|\mathcal{Z}\times|\mathcal{U}|}, \end{aligned}$$

which has rank $|\mathcal{U}|$, thus implying, by equation (23), that rank $\left(P^b(U_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}\right)=|U|$, and so equation (21) implies that

$$P^b(U_{i+1},z_i|a_i,U_i)=P^b(U_{i+1},z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}\left(P^b(U_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}\right)^{-1}. \tag{24}$$

Now, notice that equation (22) implies that

$$M_{i,a_i}P^b(Z_i|a_i,U_i)P^b(U_i|a_i,\mathcal{H}^o_{i-1})=M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1}). \tag{25}$$

Similarly notice that rank $\left(M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1})\right)=|U|$ as, from the definition of M_{i,a_i}, we have $M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1})=I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i}V^T_{i,a_i}$. This, in turn, implies that rank $\left(M_{i,a_i}P^b(Z_i|a_i,U_i)\right)=|U|$ by equation (25), and so equation (25) implies that

$$P^b(U_i|a_i,\mathcal{H}^o_{i-1})=\left(M_{i,a_i}P^b(Z_i|a_i,U_i)\right)^{-1}M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1}),$$

implying that

$$P^b(U_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}=\left(M_{i,a_i}P^b(Z_i|a_i,U_i)\right)^{-1}M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}. \tag{26}$$

Now, we claim that $M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}=I_{|\mathcal{U}|\times|\mathcal{U}|}$ since

$$\begin{aligned}M_{i,a_i}P^b(Z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}&=I_{|\mathcal{U}|\times|\mathcal{Z}|}U^T_{i,a_i}U_{i,a_i}\Sigma_{i,a_i}V^T_{i,a_i}\left(I_{|\mathcal{U}|\times|\mathcal{Z}|}U^T_{i,a_i}U_{i,a_i}\Sigma_{i,a_i}V^T_{i,a_i}\right)^+\\&=I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i}(I_{|\mathcal{U}|\times|\mathcal{Z}|}\Sigma_{i,a_i})^+\end{aligned}$$

and thus equation (26) becomes

$$P^b(U_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}=\left(M_{i,a_i}P^b(Z_i|a_i,U_i)\right)^{-1}. \tag{27}$$

Finally, substituting (27) into (24) gives

$$\begin{aligned}&P^b(U_{i+1},z_i|a_i,U_i)\\&=P^b(U_{i+1},z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}M_{i,a_i}P^b(Z_i|a_i,U_i).\end{aligned}$$

This, in combination with the fact that

$$P^b(Z_i|a_i,U_i)P^b(U_i,z_{i-1}|a_{i-1},\mathcal{H}^o_{i-2})=P^b(Z_i,z_{i-1}|a_{i-1},\mathcal{H}^o_{i-2}) \tag{28}$$

immediately implies that

$$\begin{aligned}P^b(U_{i+1},z_i|a_i,U_i)P^b(U_i,z_{i-1}|a_i,U_i)&=P(U_{i+1},z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}\\&\cdot M_{i,a_i}P^b(Z_i,z_{i-1}|a_{i-1},\mathcal{H}^o_{i-2})M'_{a_{i-1}}M_{a_{i-1}}P^b(Z_{i-1}|a_{i-1}U_{i-1}),\end{aligned}$$

as desired. Replacing U_{i+1} with r_i and i with gives the second part of the Lemma. □

Again, the proof of Theorem 3 follows first from writing $P(r_t,z_t,\dots,z_0||a_0,\dots,a_t)$ as

$$P^b(r_t,z_t|a_t,U_t)\left(\prod_{i=t-1}^{0}P^b(U_{i+1},z_i|a_i,U_i)\right)P^b(U_0),$$

using Lemma 2 and equation (28) above to inductively show that

$$\begin{aligned}&P^b(r_t,z_t|a_t,U_t)\left(\prod_{i=t-1}^{0}P^b(U_{i+1},z_i|a_i,U_i)\right)P^b(U_0)\\&\qquad=P^b(r_t,z_t|a_t,\mathcal{H}^o_{t-1})M'_{t,a_t}\prod_{i=t-1}^{0}\left(M_{i+1,a_{i+1}}P^b(Z_{i+1},z_i|a_i,\mathcal{H}^o_{i-1})M'_{i,a_i}\right)M_{0,a_0}P^b(Z_0),\end{aligned} \tag{29}$$

and again using the fact that

$$P^e(r_t)=\sum_{\tau^o\in\mathcal{T}^o_t}\Pi_e(\tau^o)P(r_t,z_t,\dots,z_0||a_0,\dots,a_t)$$

to obtain the final result.

A.4. Multi-Step Futures

While we showed that our approach cannot use extended futures as proxy for a single confounder, we now show that it is possible to incorporate *some* extended futures under prohibitive rank assumptions. In particular, we make the following assumption, in addition to Assumption 4:

Assumption 9. $\text{rank}(P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})) \geq |\mathcal{U}|^\ell, \forall i, a_i : \ell - 1 \leq i \leq H, a_{i+\ell-1} \in \mathcal{A}$.

We then obtain the following estimator:

Theorem 6. *Let $\ell > 1$. Under Assumptions 9 and 4, take $M_{i+\ell,a_{i+\ell}}$ to be the matrix of left singular vectors of $P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$ and $M'_{i+\ell,a_{i+\ell}} := (M_{i,a_i}P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}))^+$. Then the probability $P^e(r_t)$ is equal to*

$$\sum_{\tau^o \in \mathcal{T}^o_t} \Pi_e(\tau^o) P^b(r_t, z_{t-\ell+1:t}|a_t, A_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell}) M'_{t,a_t} \cdot \prod_{i=t-\ell}^{0} \Big(M_{i+\ell,a_{i+\ell}} P^b(Z_{i+1:i+\ell}, z_i|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) \cdot M'_{i+\ell-1,a_{i+\ell-1}} \Big) M_{\ell-1,a_{\ell-1}} P^b(Z_{0:\ell-1}||a_{0:\ell-2}) \quad (30)$$

While the vector $P^b(Z_{0:\ell-1}||a_{0:\ell-2})$ is not directly observable from data, it can be calculated via marginalizing out $r_{\ell-1}$ in $P^b(r_{\ell-1}, z_{\ell-1}, \ldots, z_0||a_0, \ldots, a_{\ell-1})$ and is thus estimable under Assumption 3, for all $i \leq \ell - 1$.

To prove Theorem 6, we use $z_{i:i+\ell-1}$ and $(h^o_{i-1}, a_{i:i+\ell-1})$ as proxy variables for the confounder $u_{i:i+\ell-1}$ under the treatment $a_{i+\ell-1}$. As seen in Figure **??**, the causal structure is the same as that in both the one-step proxy and extended history cases, allowing for a similar analysis, which we now present.

A.4.1. Necessity of Assumption 4

First, we show that Assumption 4 is a necessary condition for Assumption 9. This is simply because

$$P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$

can be written as

$$P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, U_i \times A_{i:i+\ell-2}) P^b(U_i \times A_{i:i+\ell-2}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$

so that

$$\text{rank}(P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})) \leq |\mathcal{U}||\mathcal{A}|^{\ell-1}.$$

Thus, Assumption 9 cannot be met if $|\mathcal{A}| < |\mathcal{U}|$, hence the necessity of Assumption 4.

A.4.2. Proof of Theorem 6

Again, the proof of Theorem 6 is essentially the same as those for Theorems 2 and 3 except that we decompose the probability $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$ in a slightly different way. We proceed, again, by proving the analogous Lemma to Lemmas 1 and 2:

Lemma 3. *Under Assumptions 9 and 4, take $M_{i+\ell,a_{i+\ell}}$ to be the matrix of left singular vectors of $P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$ and $M'_{i+\ell,a_{i+\ell}} := (M_{i,a_i}P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}))^+$. Then*

$$\begin{aligned} &P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1}) P^b(U_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, U_{i-1:i+\ell-2}) \\ &\quad = P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} M_{i+\ell-1,a_{i+\ell-1}} \\ &\quad\quad \cdot P^b(Z_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, \mathcal{A}_{i-1:i+\ell-3} \times \mathcal{H}^o_{i-2}) M'_{i+\ell-2,a_{i+\ell-2}} \\ &\quad\quad \cdot M_{i+\ell-2,a_{i+\ell-2}} P^b(Z_{i-1:i+\ell-2}|U_{i-1:i+\ell-2}) \end{aligned}$$

and

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t})P^b(U_{t-\ell+1:t}, z_{t-\ell}|a_{t-1}, U_{t-\ell:t-1})$$
$$= P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}M_{t,a_t}P^b(Z_{t-\ell+1:t}, z_{t-\ell}|a_{t-1}, \mathcal{A}_{t-\ell:t-2} \times \mathcal{H}^o_{t-\ell-1})$$
$$\cdot M'_{t-1,a_{t-1}}M_{t-1,a_{t-1}}P^b(Z_{t-\ell:t-1}|U_{t-\ell:t-1})$$

Proof. First, notice that

$$P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1})P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$
$$= P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$

$$\implies P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1})P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}$$
$$= P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}. \quad (31)$$

Additionally,

$$P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) = P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1})P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) \quad (32)$$

$$\implies P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}$$
$$= P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1})P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}. \quad (33)$$

From the definition of $M'_{i+\ell-1,a_{i+\ell-1}}$, we have that

$$P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}$$

$$= P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$
$$\cdot (M_{i+\ell-1,a_{i+\ell-1}}P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}))^+$$
$$= P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})V_{i+\ell-1,a_{i+\ell-1}}(I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell}\Sigma_{i+\ell-1,a_{i+\ell-1}})^+$$
$$= U_{i+\ell-1,a_{i+\ell-1}}\Sigma_{i+\ell-1,a_{i+\ell-1}}(I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell}\Sigma_{i+\ell-1,a_{i+\ell-1}})^+ = U_{i+\ell-1,a_{i+\ell-1}}I_{|\mathcal{Z}|^\ell \times |\mathcal{U}|^\ell},$$

which has rank $|\mathcal{U}|^\ell$, thus implying, by equation (33), that

$$\text{rank}\left(P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}\right) = |U|^\ell,$$

and so equation (31) implies that

$$P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1})$$
$$= P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}}$$
$$\cdot (P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}})^{-1}. \quad (34)$$

Now, notice that equation (32) implies that

$$M_{i+\ell-1,a_{i+\ell-1}}P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1})P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})$$
$$= M_{i+\ell-1,a_{i+\ell-1}}P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}). \quad (35)$$

Similarly notice that

$$\text{rank}\left(M_{i+\ell-1,a_{i+\ell-1}}P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})\right) = |U|^\ell$$

as, from the definition of $M_{i+\ell-1,a_{i+\ell-1}}$, we have

$$M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) \\ = I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell} \Sigma_{i+\ell-1,a_{i+\ell-1}} V^T_{i+\ell-1,a_{i+\ell-1}}.$$

This, in turn, implies that rank $\left(M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1})\right) = |U|^\ell$ by equation (35), and so equation (35) implies that

$$P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) \\ = (M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1}))^{-1} M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}),$$

implying that

$$P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} \\ = (M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1}))^{-1} M_{i+\ell-1,a_{i+\ell-1}} \\ \cdot P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}}. \quad (36)$$

Now, we claim that $M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} = I_{|\mathcal{U}|^\ell \times |\mathcal{U}|^\ell}$ since

$$M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}}$$

$$= I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell} U^T_{i+\ell-1,a_{i+\ell-1}} U_{i+\ell-1,a_{i+\ell-1}} \Sigma_{i+\ell-1,a_{i+\ell-1}} V^T_{i+\ell-1,a_{i+\ell-1}} \\ \cdot \left(I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell} U^T_{i+\ell-1,a_{i+\ell-1}} U_{i+\ell-1,a_{i+\ell-1}} \Sigma_{i+\ell-1,a_{i+\ell-1}} V^T_{i+\ell-1,a_{i+\ell-1}}\right)^+ \\ = I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell} \Sigma_{i+\ell-1,a_{i+\ell-1}} (I_{|\mathcal{U}|^\ell \times |\mathcal{Z}|^\ell} \Sigma_{i+\ell-1,a_{i+\ell-1}})^+$$

and thus equation (36) becomes

$$P^b(U_{i:i+\ell-1}|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} \\ = (M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1}))^{-1}. \quad (37)$$

Finally, substituting (37) into (34) gives

$$P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1}) \\ = P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} \\ \cdot M_{i+\ell-1,a_{i+\ell-1}} P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1}). \quad (38)$$

This, in combination with the fact that

$$P^b(Z_{i:i+\ell-1}|U_{i:i+\ell-1}) P^b(U_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, \mathcal{A}_{i-1:i+\ell-3} \times \mathcal{H}^o_{i-2}) \\ = P^b(Z_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, \mathcal{A}_{i-1:i+\ell-3} \times \mathcal{H}^o_{i-2}) \quad (39)$$

immediately implies that

$$P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1}) P^b(U_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, U_{i-1:i+\ell-2}) \\ = P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, \mathcal{A}_{i:i+\ell-2} \times \mathcal{H}^o_{i-1}) M'_{i+\ell-1,a_{i+\ell-1}} M_{i+\ell-1,a_{i+\ell-1}} \\ \cdot P^b(Z_{i:i+\ell-1}, z_{i-1}|a_{i+\ell-2}, \mathcal{A}_{i-1:i+\ell-3} \times \mathcal{H}^o_{i-2}) M'_{i+\ell-2,a_{i+\ell-2}} \\ \cdot M_{i+\ell-2,a_{i+\ell-2}} P^b(Z_{i-1:i+\ell-2}|U_{i-1:i+\ell-2})$$

as desired.

For the second identity, we first note that

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t})P^b(U_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell}) \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})$$

$$\implies P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t})P^b(U_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t} \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}, \quad (40)$$

which, by the fact that $\text{rank}(P^b(U_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}) = |\mathcal{U}|^\ell$, as shown above, implies that

$$\implies P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t}) \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}(P^b(U_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t})^{-1}. \quad (41)$$

Substituting equation (37) with $i = t - \ell + 1$ into equation (41), then gives

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t}) \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}M_{t,a_t}P^b(Z_{t-\ell+1:t}|U_{t-\ell+1:t}),$$

implying, via equations (39) and (38) with $i = t - \ell + 1$ and $i = t - \ell$, respectively, that

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t})P^b(U_{t-\ell+1:t}, z_{t-\ell}|a_{t-1}, U_{t-\ell:t-1}) \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, \mathcal{A}_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t}M_{t,a_t}P^b(Z_{t-\ell+1:t}, z_{t-\ell}|a_{t-1}, \mathcal{A}_{t-\ell:t-2} \times \mathcal{H}^o_{t-\ell-1}) \\ \cdot M'_{t-1,a_{t-1}}M_{t-1,a_{t-1}}P^b(Z_{t-\ell:t-1}|U_{t-\ell:t-1}),$$

as desired. ☐

Finally, the proof of Theorem 6 follows first from writing $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$ in the following, slightly different way from the previous two sections:

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t}) \left(\prod_{i=t-\ell}^{0} P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1}) \right) P^b(U_{0:\ell-1}||a_{0:\ell-2}).$$

The result then follows from using Lemma 3 and equation (39) above to inductively show that

$$P^b(r_t, z_{t-\ell+1:t}|a_t, U_{t-\ell+1:t}) \left(\prod_{i=t-\ell}^{0} P^b(U_{i+1:i+\ell}, z_i|a_{i+\ell-1}, U_{i:i+\ell-1}) \right) P^b(U_{0:\ell-1}||a_{0:\ell-2}) \\ = P^b(r_t, z_{t-\ell+1:t}|a_t, A_{t-\ell+1:t-1} \times \mathcal{H}^o_{t-\ell})M'_{t,a_t} \prod_{i=t-\ell}^{0} \Big(M_{i+\ell,a_{i+\ell}} \\ \cdot P^b(Z_{i+1:i+\ell}, z_i|a_{i+\ell-1}, A_{i:i+\ell-2} \times \mathcal{H}^o_{i-1})M'_{i+\ell-1,a_{i+\ell-1}} \Big) M_{\ell-1,a_{\ell-1}}P^b(Z_{0:\ell-1}||a_{0:\ell-2}),$$

and again using the fact that

$$P^e(r_t) = \sum_{\tau^o \in \mathcal{T}^o_t} \Pi_e(\tau^o)P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$$

to obtain the final result.

A.4.3. Identification of $P^b(Z_{0:\ell-1}||a_{0:\ell-2})$

As mentioned in Section 5, the identification of $P^b(Z_{0:\ell-1}||a_{0:\ell-2})$ in Theorem 6 can be achieved via the estimator in Theorem 3. In particular, as mentioned in Section A.3.1, the estimator rests on the identification of $P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t)$ as

$$P^b(r_t, z_t|a_t, \mathcal{H}^o_{t-1})\tilde{M}'_{t,a_t} \prod_{i=t-1}^{0} \left(\tilde{M}_{i+1,a_{i+1}} P^b(Z_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})\tilde{M}'_{i,a_i}\right) \tilde{M}_{0,a_0} P^b(Z_0),$$

under Assumption 3, and where $\tilde{M}_{i,a_i}$ is the matrix of left singular vectors of $P^b(Z_i|a_i, \mathcal{H}^o_{i-1})$ and $\tilde{M}'_{i,a_i} := (M_{i,a_i} P^b(Z_i|a_i, \mathcal{H}^o_{i-1}))^+$. Hence, marginalizing out r_t, we see have the following

Corollary 6.1. *Under Assumption 3, we have that*

$$\begin{aligned} &P(r_t, z_t, \ldots, z_0||a_0, \ldots, a_t) \\ &\quad = P^b(z_t|a_t, \mathcal{H}^o_{t-1})\tilde{M}'_{t,a_t} \prod_{i=t-1}^{0} \left(\tilde{M}_{i+1,a_{i+1}} P^b(Z_{i+1}, z_i|a_i, \mathcal{H}^o_{i-1})\tilde{M}'_{i,a_i}\right) \tilde{M}_{0,a_0} P^b(Z_0) \end{aligned}$$

Hence, Corollary 6.1 allows for the identification of $P^b(Z_{0:\ell-1}||a_{0:\ell-2})$ under Assumption 3.

B. OPE via Importance Sampling

We now give a derivation for the IS estimator given in Section 6. We have that

$$\begin{aligned} v_H(\pi_e) &= \sum_{\tau \in \mathcal{T}_H} \sum_v v P^e(v|\tau) P^e(\tau) \\ &= \sum_{\tau \in \mathcal{T}_H} \sum_v v P^b(v|u_{0:H}) P^b(\tau) \left(\frac{\Pi_e(\tau^o)}{\Pi_b(u_{0:H})}\right) \\ &= \sum_v v \sum_{\tau^o \in \mathcal{T}^o_t} P^b(\tau^o)\Pi_e(\tau^o) \sum_{u_{0:H}} \frac{P^b(v|u_{0:H})P^b(u_{0:H}|\tau^o)}{\Pi_b(u_{0:H})} \\ &= \sum_v v \sum_{\tau^o \in \mathcal{T}^o_t} P^b(v|\tau^o) W_{e,b}(v, \tau^o) = \mathbb{E}_{\tau^o}\left[\mathbb{E}_v[v W_{e,b}(v, \tau^o)|\tau^o]\right], \end{aligned}$$

where the second and penultimate equalities are made possible by Assumption 8.

Thus, we can write

$$v_H(\pi_e) = \mathbb{E}_{\tau^o}\left[\mathbb{E}_v[v W(v, \tau^o)|\tau^o]\right] \tag{42}$$

with

$$W(v, \tau^o) := \frac{P^b(\tau^o)}{P^b(v|\tau^o)} \Pi_e(\tau^o)\Gamma_b(v, \tau^o) \tag{43}$$

and

$$\Gamma_b(v, \tau^o) = \sum_{r_{0:H}: \sum r_i = v} \sum_{u_{0:H}} \frac{\prod_{i=0}^H P^b(r_i|u_i)P^b(u_{0:H}|\tau^o)}{\Pi_b(u_{0:H})}. \tag{44}$$

Hence, all that must be shown is the identifiability of $\Gamma_b(v, \tau^o)$, as the terms of $\frac{P^b(\tau^o)}{P^b(v|\tau^o)}\Pi_e(\tau^o)$ are all either given or directly estimable from observable data.

B.1. Identifiability of $\Gamma_b(v, \tau^o)$

First, we show that under certain rank and distinctness assumptions, the vectors $\pi_b^{(i)}(a_i|U_i)$ and $P^b(r_i|U_i)$ are identifiable by extending the diagonalization method of (Kuroki & Pearl, 2014). We will treat both of the above vectors individually. However, letting $\kappa = |\mathcal{U}|$, we first define a matrix which will be used in both derivations:

$$\Delta_{i,z_{i+1}} := \text{diag}(P^b(z_{i+1}|u_i^{(1)}), \ldots, P^b(z_{i+1}|u_i^{(\kappa)}).$$

The ordering $u_i^{(j)}$ is such that the elements on the diagonal of $\Delta_{i,z_{i+1}}$ are in non-decreasing order. That is, $P^b(z_{i+1}|u_i^{(1)}) \geq \ldots \geq P^b(z_{i+1}|u_i^{(\kappa)}))$. Furthermore, we will let z^j, r^j, and a^j denote the jth elements of the sets $\mathcal{Z}$, $\mathcal{R}$, and $\mathcal{A}$, respectively. Henceforth, the vectors $\pi_b^{(i)}(a_i|U_i)$ and $P^b(z_i|U_i)$ are defined such that their jth entries are $\pi_b^{(i)}(a_i|u_i^{(j)})$ and $P^b(z_i|u_i^{(j)})$, respectively. Throughout our identification procedure for both of these vectors, we make the following crucial distinctness assumption, which will allow for unique eigendecomposition (up to constants):

Assumption 10 (Distinctness)**.** *For each $z_{i+1} \in \mathcal{Z}$, the probabilities $\{P^b(z_{i+1}|u_i) : u_i \in \mathcal{U}\}$ are all distinct.*

B.1.1. Identifying $\pi_b^{(i)}(a_i|U_i)$

Similar to those defined by (Kuroki & Pearl, 2014), we first define the following matrices which will be key to the analysis, explicitly state all assumptions in terms of them, and then show the result.

Define the matrix following matrices

$$P^a_{i,z_i} := \begin{pmatrix} 1 & P^b(a_i^1|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}|z_i) \\ P^b(z_{i-1}^1|z_i) & P^b(a_i^1, z_{i-1}^1|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}, z_{i-1}^1|z_i) \\ \vdots & \vdots & \ddots & \vdots \\ P^b(z_{i-1}^{|\mathcal{Z}|-1}|z_i) & P^b(z_{i-1}^{|\mathcal{Z}|-1}, a_i^1|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}, z_{i-1}^{|\mathcal{Z}|-1}|z_i) \end{pmatrix},$$

$$Q^a_{i,z_i,z_{i+1}} := \begin{pmatrix} P^b(z_{i+1}|z_i) & P^b(a_i^1, z_{i+1}|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}, z_{i+1}|z_i) \\ P^b(z_{i-1}^1, z_{i+1}|z_i) & P^b(a_i^1, z_{i-1}^1, z_{i+1}|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}, z_{i-1}^1, z_{i+1}|z_i) \\ \vdots & \vdots & \ddots & \vdots \\ P^b(z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}|z_i) & P^b(a_i^1, z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}|z_i) & \cdots & P^b(a_i^{|\mathcal{A}|-1}, z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}|z_i) \end{pmatrix},$$

$$U^a_i := \begin{pmatrix} 1 & \pi_b^{(i)}(a_i^1|u_i^{(1)}) & \cdots & \pi_b^{(i)}(a_i^{|\mathcal{A}|-1}|u_i^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \pi_b^{(i)}(a_i^1|u_i^{(\kappa)}) & \cdots & \pi_b^{(i)}(a_i^{|\mathcal{A}|-1}|u_i^{(\kappa)}) \end{pmatrix},$$

$$R^a_{i,z_i} = \begin{pmatrix} 1 & P^b(z_{i-1}^1|z_i, u_i^{(1)}) & \cdots & P^b(z_{i-1}^{|\mathcal{Z}|-1}|z_i, u_i^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & P^b(z_{i-1}^1|z_i, u_i^{(\kappa)}) & \cdots & P^b(z_{i-1}^{|\mathcal{Z}|-1}|z_i, u_i^{(\kappa)}) \end{pmatrix},$$

$$M^a_{i,z_i} := \text{diag}(P^b(u_i^{(1)}|z_i), \ldots, P^b(u_i^{(\kappa)}|z_i)).$$

In addition to Assumptions 4, 5, 6, and 10, we also make the following rank assumption:

Assumption 11 (Rank)**.** *The matrices P^a_{i,z_i} and Q^a_{i,z_i} have rank at least $|\mathcal{U}|$.*

With these conditions, we now show how to obtain the matrix U_i^a, which contains the desired probabilities, via diagonalization of observable matrices.

Notice that

$$P^a_{i,z_i} = {R^a_{i,z_i}}^T M^a_{i,z_i} U^a_i \tag{45}$$

and

$$Q^a_{i,z_i,z_{i+1}} = {R^a_{i,z_i}}^T M^a_{i,z_i} \Delta_{i,z_{i+1}} U^a_i \tag{46}$$

Now, for any subset Λ of $\{2, \ldots, |\mathcal{Z}|\}$ of size $|\mathcal{U}| - 1$, define $U_i^a(\Lambda)$ to be the $|\mathcal{U}| \times |\mathcal{U}|$ matrix whose first column is the first column of U_i^a and whose jth column is the Λ_jth column of U_i^a for all $2 \leq j \leq |\mathcal{U}|$, where Λ_j is the jth smallest element of Λ. Equivalently, $U_i^a(\Lambda)$ can be written as $U_i^a L(\Lambda)$ where $L(\Lambda)$ is the $|\mathcal{Z}| \times |\mathcal{U}|$ matrix whose first column is e_1 and whose jth column is e_{Λ_j}, where e_k denotes the kth standard basis vector of $\mathbb{R}^{|\mathcal{Z}|}$, for $2 \leq j \leq |\mathcal{U}|$. Right-multiplying equations (45) and (46) by $L(\Lambda)$ thus give, for each such subset Λ,

$$P^a_{i,z_i} L(\Lambda) = {R^a_{i,z_i}}^T M^a_{i,z_i} U^a_i(\Lambda) \tag{47}$$

and

$$Q^a_{i,z_i,z_{i+1}} L(\Lambda) = {R^a_{i,z_i}}^T M^a_{i,z_i} \Delta_{i,z_{i+1}} U^a_i(\Lambda) \tag{48}$$

Notice that Assumption 11 implies that $\text{rank}(U^a_i) = |\mathcal{U}|$ via equation (45), and thus that $U^a_i(\Lambda)$ is invertible for all such sets Λ. Additionally, Assumption 11 implies that M^a_{i,z_i} is invertible and $\text{rank}(R^a_{i,z_i}) = |\mathcal{U}|$, via equation (45), and so, from equation (47), we have $(P^a_{i,z_i} L(\Lambda))^+ = (U^a_i(\Lambda))^{-1}(M^a_{i,z_i})^{-1}({R^a_{i,z_i}}^T)^+$. This along with equation (48) gives the following:

Lemma 4. *For any $\Lambda \subset \{2, \dots, |\mathcal{Z}|\}$ of size $|\mathcal{U}| - 1$, under Assumption 11, we have that*

$$(P^a_{i,z_i} L(\Lambda))^+ Q^a_{i,z_i,z_{i+1}} L(\Lambda) = (U^a_i(\Lambda))^{-1} \Delta_{i,z_{i+1}} U^a_i(\Lambda).$$

We now show how to identify $U^a_i(\Lambda)$ for any such subset Λ.

Lemma 5. *The matrices $U^a_i(\Lambda)$ are identifiable from observable matrices.*

Proof. We use the diagonalization strategy of (Kuroki & Pearl, 2014). At a high-level, the procedure is as follows:

1. Diagonalize the observable matrix $(P^a_{i,z_i} L(\Lambda))^+ Q^a_{i,z_i,z_{i+1}} L(\Lambda)$ with the eigenvalues of the diagonal matrix in descending order to get $T_{i,z_i,z_{i+1}}(\Lambda) \Delta_{i,z_{i+1}} T^{-1}_{i,z_i,z_{i+1}}(\Lambda)$
2. Recover $U^a_i(\Lambda)$ from $T_{i,z_i,z_{i+1}}(\Lambda)$.

Notice, importantly, that in the first step of the above, the matrix of eigenvalues is $\Delta_{i,z_{i+1}}$. This is guaranteed because Lemma 4 writes the matrix $(P^a_{i,z_i} L(\Lambda))^+ Q^a_{i,z_i,z_{i+1}} L(\Lambda)$ in diagonal form, where we have specified the ordering such that the entries of the diagonal matrix are in decreasing order. Furthermore, the distinctness of the diagonal entries of $\Delta_{i,z_{i+1}}$ assumed in Assumption 10 ensures uniqueness of the eigenvectors up to constant factors, thus implying that the columns of $(U^a_i(\Lambda))^{-1}$ are constant multiples of the columns of the observable matrix $T_{i,z_i,z_{i+1}}(\Lambda)$. We show how these are recovered in the second step.

Let $C_{T_{i,z_i,z_{i+1}}(\Lambda)} = \text{diag}(c^1_{i,z_i,z_{i+1}}, \dots, c^\kappa_{i,z_i,z_{i+1}})$ be the diagonal matrix of constants such that

$$(U^a_i(\Lambda))^{-1} = T_{i,z_i,z_{i+1}}(\Lambda) C_{T_{i,z_i,z_{i+1}}(\Lambda)}. \tag{49}$$

As the left hand side is invertible, the right hand side is also invertible, and so equation (49) implies that

$$U^a_i(\Lambda) = C^{-1}_{T_{i,z_i,z_{i+1}}(\Lambda)} T^{-1}_{i,z_i,z_{i+1}}(\Lambda).$$

Finally, as the first column of $U^a_i(\Lambda)$ is $\vec{1}$, the matrix $C^{-1}_{T_{i,z_i,z_{i+1}}(\Lambda)}$ can be written in terms of $T^{-1}_{i,z_i,z_{i+1}}(\Lambda)$, and consequently, so can its inverse $C_{T_{i,z_i,z_{i+1}}(\Lambda)}$, thus allowing for the identification of $U^a_i(\Lambda) = T_{i,z_i,z_{i+1}}(\Lambda) C_{T_{i,z_i,z_{i+1}}(\Lambda)}$. □

The above allows for the columns given the set Λ of U^a_i to be identified. Hence, choosing multiple different such Λ's of size $|\mathcal{U}| - 1$, such that their union is $\{2, \dots, |\mathcal{Z}|\}$ (i.e., this will require $\lceil \frac{|\mathcal{Z}|-1}{|\mathcal{U}|-1} \rceil$ different choices of Λ), the entire matrix U^a_i can be identified, and thus all probabilities $\pi^{(i)}_b(a_i | u^{(j)}_i)$ can be written in terms of observables, as desired.

B.1.2. Identifying $P^b(r_i | U_i)$

The identification procedure for $P^b(r_i | U_i)$ is essentially the same as that in the previous section, except that we use the following matrices:

$$P^r_{i,a_i,z_i} = \begin{pmatrix} 1 & P^b(r^1_i | z_i) & \cdots & P^b(r^{|\mathcal{R}|-1}_i | z_i) \\ P^b(z^1_{i-1}, a_i | z_i) & P^b(r^1_i, z^1_{i-1}, a_i | z_i) & \cdots & P^b(r^{|\mathcal{R}|-1}_i, z^1_{i-1}, a_i | z_i) \\ \vdots & \vdots & \ddots & \vdots \\ P^b(z^{|\mathcal{Z}|-1}_{i-1}, a_i | z_i) & P^b(r^1_i, z^{|\mathcal{Z}|-1}_{i-1}, a_i | z_i) & \cdots & P^b(r^{|\mathcal{R}|-1}_i, z^{|\mathcal{Z}|-1}_{i-1}, a_i | z_i) \end{pmatrix},$$

$$Q^r_{i,z_i,z_{i+1},a_i} = \begin{pmatrix} P^b(z_{i+1}|z_i) & P^b(r_i^1, z_{i+1}|z_i) & \cdots & P^b(r_i^{|\mathcal{R}|-1}, z_{i+1}|z_i) \\ P^b(z_{i-1}^1, z_{i+1}, a_i|z_i) & P^b(r_i^1, z_{i-1}^1, z_{i+1}, a_i|z_i) & \cdots & P^b(r_i^{|\mathcal{R}|-1}, z_{i-1}^1, z_{i+1}, a_i|z_i) \\ \vdots & \vdots & \ddots & \vdots \\ P^b(z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}, a_i|z_i) & P^b(r_i^1, z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}, a_i|z_i) & \cdots & P^b(r_i^{|\mathcal{R}|-1}, z_{i-1}^{|\mathcal{Z}|-1}, z_{i+1}, a_i|z_i) \end{pmatrix},$$

$$U_i^r = \begin{pmatrix} 1 & P^b(r_i^1|u_i^{(1)}) & \cdots & P^b(r_i^{|\mathcal{R}|-1}|u_i^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & P^b(r_i^1|u_i^{(k)}) & \cdots & P^b(r_i^{|\mathcal{R}|-1}|u_i^{(k)}) \end{pmatrix},$$

$$R^r_{i,z_i,a_i} = \begin{pmatrix} 1 & P^b(a_i, z_{i-1}^1|z_i, u_i^{(1)}) & \cdots & P^b(a_i, z_{i-1}^{|\mathcal{Z}|-1}|z_i, u_i^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & P^b(a_i, z_{i-1}^1|z_i, u_i^{(k)}) & \cdots & P^b(a_i, z_{i-1}^{|\mathcal{Z}|-1}|z_i, u_i^{(k)}) \end{pmatrix},$$

and

$$M^r_{i,z_i} = \text{diag}(P^b(u_i^{(1)}|z_i), \ldots, P^b(u_i^{(k)}|z_i)).$$

Again, analogously to the previous section, in addition to Assumptions 4, 5, 6, and 10, we also make the following rank assumption:

Assumption 12 (Rank)**.** *The matrices P^r_{i,z_i} and Q^r_{i,z_i} have rank at least $|\mathcal{U}|$.*

Importantly, we have the same identities on the above matrices:

$$P^r_{i,z_i} = {R^r_{i,z_i}}^T M^r_{i,z_i} U_i^a \tag{50}$$

and

$$Q^r_{i,z_i,z_{i+1}} = {R^r_{i,z_i}}^T M^r_{i,z_i} \Delta_{i,z_{i+1}} U_i^r \tag{51}$$

The procedure for identifying U_i^r is then precisely the same as that described in the previous section. Importantly, because the diagonal matrix is the *same* matrix $\Delta_{i,z_{i+1}}$ used to identify $\pi_b^{(i)}(a_i|U_i)$, the ordering $u_i^{(1)}, \ldots, u_i^{(k)}$ is the same, so that when we identify the vector $P^b(r_i|U_i)$, the ordering of entries is indeed in corresponding order to our identification of $\pi_b^{(i)}(a_i|U_i)$.

We summarize the results from the previous 2 sections here before we proceed with the rest of the identification procedure for $\Gamma_b(v, \tau^o)$.

Theorem 7. *Under Assumptions 4, 5, 6, 10, 11, and 12, the vectors $\pi_b^{(i)}(a_i|U_i)$ and $P^b(r_i|U_i)$ are identifiable, under the consistent ordering that their jth entries are $\pi_b^{(i)}(a_i|u_i^{(j)})$ and $P^b(r_i|u_i^{(j)})$, respectively.*

We now proceed with the identification procedure for $\Gamma_b(v, \tau^o)$. The sole remaining ingredient is the identification of $P^b(U_{0:H}|\mathcal{T}_H^o)$, which we now derive.

B.1.3. Identifying $P^b(U_{0:H}|\mathcal{T}_H^o)$

We now propose an identification scheme for $P^b(U_{0:H}|\mathcal{T}_H^o)$ which rests on a Bayes' rule and eigenvalue analysis argument. Our two key assumptions will be an exponential rank assumption in H and an additional probability positivity assumption:

Assumption 13 (Rank)**.** $\text{rank}(P^b(U_{0:H}|\mathcal{T}_H^o)) = |\mathcal{U}|^{H+1}$.

Assumption 14 (Positivity)**.** $P^b(\tau^o) > 0, \forall \tau^o \in \mathcal{T}_H^o$.

With these two assumptions we now derive the identification procedure for $P^b(U_{0:H}|\mathcal{T}_H^o)$.

By Bayes' Rule, we have that

$$\begin{aligned} &P^b(a_0, \ldots, a_H|u_0, \ldots, u_H, z_0, \ldots, z_H) \\ &= \frac{P^b(u_0, \ldots, u_H|\tau^o)P^b(\tau^o)}{P^b(u_0, \ldots, u_H|\mathcal{T}_H^o)P^b(\mathcal{T}_H^o)}. \end{aligned}$$

Now, let $P^b_{u_0,\ldots,u_H}$ denote the column vector in $\mathbb{R}^{(|\mathcal{A}||\mathcal{Z}|)^{H+1}}$ whose jth entry is $P^b(a_0,\ldots,a_H|u_0,\ldots,u_H,z_0,\ldots,z_H)$, where $\tau^o=(z_0,a_0,\ldots,z_H,a_H)$ is the jth element of the set $\mathcal{T}^o_H$. Then, letting $\frac{1}{P^b(\mathcal{T}^o_H)}$ denote the vector whose jth element is $\frac{1}{(P^b(\mathcal{T}^o_H))_j}$ (which is allowable under Assumption 14) and letting $\odot$ denote Hadamard product, we have that

$$P^b_{u_0,\ldots,u_H}\odot\frac{1}{P^b(\mathcal{T}^o_H)}=\frac{P^b(u_0,\ldots,u_H|\mathcal{T}^o_H)^T}{P^b(u_0,\ldots,u_H|\mathcal{T}^o_H)P^b(\mathcal{T}^o_H)},$$

implying, for each $(j_0,\ldots,j_H)\in[\kappa]^{H+1}$, that

$$P^b_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}\odot\frac{1}{P^b(\mathcal{T}^o_H)}=\frac{P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T}{P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)P^b(\mathcal{T}^o_H)}. \tag{52}$$

Now, we claim that the left hand side of the above is identifiable. This is because, firstly, the vector $P^b(\mathcal{T}^o_H)$ is clearly directly estimable from observable data and, secondly, each entry in the vector $P^b_{u_0,\ldots,u_H}$ is of the form

$$P^b(a_0,\ldots,a_H|u_0^{(j_0)},\ldots,u_H^{(j_H)},z_0,\ldots,z_H)=\prod_{i=0}^{H}\pi_b^{(i)}(a_i|u_i^{(j_i)}),$$

and we have shown the indentification procedure for $\pi_b^{(i)}(a_i|U_i)$ in Section B.1.1. Then equation (52) implies that

$$\left(P^b_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}\odot\frac{1}{P^b(\mathcal{T}^o_H)}\right)P^b(\mathcal{T}^o_H)^TP^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T=P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T.$$

Defining $\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})$ to be the outer product $\left(P^b_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}\odot\frac{1}{P^b(\mathcal{T}^o_H)}\right)P^b(\mathcal{T}^o_H)^T$, we note that from what we have said, $\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})$ is also identifiable.

Thus, we must identify $P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T$ from the equation

$$\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T=P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)^T \tag{53}$$

Equation (53) implies that $\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})$ has 1 as an eigenvalue. But as $\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})$ can be written as an outer product of length $|\mathcal{T}^o_H|$ vectors, it has 0 as an eigenvalue with multiplicity $|\mathcal{T}^o_H|-1$, thus implying that the multiplicity of the eigenvalue 1 is 1.

Hence, diagonalizing the observable matrix $\Omega(u_0^{(j_0},\ldots,u_H^{(j_H)})$ allows for identification of $P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\mathcal{T}^o_H)$ up to an unknown constant factor $\alpha_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}$. Thus, all that remains to show is the recovery of $\alpha_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}$.

Letting $A_{\mathbf{j}}$ be the diagonal matrix whose entries, for each sequence $\mathbf{j}\in[\kappa]^{H+1}$, are $\alpha_{u_0^{(j_0)},\ldots,u_H^{(j_H)}}$ in order corresponding to the rows of $P^b(U_{0:H}|\mathcal{T}^o_H)$, the above identification scheme allows us to recover the matrix

$$A_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H)$$

with $A_{\mathbf{j}}$ unknown but all entries of A_j non-zero (i.e. since the rows of $A_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H)$ are eigenvectors).

Letting $b_{\mathbf{j}}\in\mathbb{R}^{|\mathcal{U}|^{H+1}}$, note that the equation

$$b_{\mathbf{j}}^TA_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H)=\vec{1}^T,$$

has a solution given by the vector $b^*_{\mathbf{j}}$ such that $(b^*_{\mathbf{j}})_k=(A_{\mathbf{j}}^{-1})_{kk}$. Note also, however, that Assumption 13, guarantees that $P^b(U_{0:H}|\mathcal{T}^o_H)$ is of full row rank, thus implying that

$$b_{\mathbf{j}}^T=\vec{1}^T(A_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H))^+.$$

Hence,

$$(A_{\mathbf{j}}^{-1})_{kk}=(\vec{1}^T(A_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H))^+)_k,$$

and so the matrix $A_{\mathbf{j}}$ is recoverable from the observable matrix $A_{\mathbf{j}}P^b(U_{0:H}|\mathcal{T}^o_H))$, thus implying that the matrix $P^b(U_{0:H}|\mathcal{T}^o_H)$ is also identifiable, as desired.

B.1.4. COMBINING IDENTIFICATIONS

Finally, recalling that

$$\Gamma_b(v,\tau^o) = \sum_{r_{0:H}:\sum r_i = v} \sum_{u_{0:H}} \frac{\prod_{i=0}^H P^b(r_i|u_i) P^b(u_{0:H}|\tau^o)}{\Pi_b(u_{0:H})},$$

$$= \sum_{r_{0:H}:\sum r_i = v} \sum_{1\le j_0,\ldots,j_H\le \kappa} \frac{\prod_{i=0}^H P^b(r_i|u_i^{(j_i)}) P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\tau^o)}{\Pi_b(u_0^{(j_0)},\ldots,u_H^{(j_H)})},$$

the above derivations allow for identification for *each* of the terms $P^b(r_i|u_i^{(j_i)})$, $P^b(u_0^{(j_0)},\ldots,u_H^{(j_H)}|\tau^o)$, and $\Pi_b(u_0^{(j_0)},\ldots,u_H^{(j_H)})$, thus implying the desired identifiability for the entire IS estimation procedure, as desired.

We summarize this result and all conditions below:

Theorem 8. *Under Assumptions 4, 5, 6, 7, 8, 10, 11, 12, 13, and 14, the quantity $\Gamma_b(v,\tau^o)$ is identifiable for all values v and $\tau^o \in \mathcal{T}_H^o$. Hence, the importance weights*

$$W(v,\tau^o) := \frac{P^b(\tau^o)}{P^b(v|\tau^o)} \Pi_e(\tau^o)\Gamma_b(v,\tau^o)$$

are all identifiable, allowing for the IS procedure given by $v_H(\pi_e) = \mathbb{E}_{\tau^o}\left[\mathbb{E}_v[vW(v,\tau^o)|\tau^o]\right]$