Semiparametric Estimation of Markov Decision Processes with Continuous State Space*

Sorawoot Srisuma† and Oliver Linton‡
London School of Economics and Political Science

The Suntory Centre
Suntory and Toyota International Centres for Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6674

Discussion paper no.:
EM/2010/550
August 2010

---

* We thank Xiaohong Chen, Philipp Schmidt-Dengler, and seminar participants at the 19th EC-Squared Conference on .Recent Advances in Structural Microeconometrics. in Rome, and Workshop in Banff Semiparametric and Nonparmetric Methods in Econometrics.in Banff, for helpful comments. This research is supported by the ESR, United Kingdom.

† Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom E-mail address: s.t.srisuma@lse.ac.uk

‡ Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk

Abstract

We propose a general two-step estimation method for the structural parameters of popular semiparametric Markovian discrete choice models that include a class of Markovian Games and
allow for continuous observable state space. The estimation procedure is simple as it directly generalizes the computationally attractive methodology of Pesendorfer and Schmidt-Dengler (2008) that assumed finite observable states. This extension is non-trivial as the value functions, to be estimated nonparametrically in the first stage, are defined recursively in a non-linear functional equation. Utilizing structural assumptions, we show how to consistently estimate the infinite dimensional parameters as the solution to some type II integral equations, the solving of which is a well-posed problem. We provide sufficient set of primitives to obtain root-T consistent estimators for the finite dimensional structural parameters and the distribution theory for the value functions in a time series framework.

Keywords: Discrete Markov Decision Models, Kernel Smoothing, Markovian Games, Semi-parametric Estimation, Well-Posed Inverse Problem.

1 Introduction

The inadequacy of static frameworks to model economic phenomena led to the development of recursive methods in economics. The mathematical theory underlying discrete time modelling is dynamic programming developed by Bellman (1957); for a review of its prevalence in modern economic theory, see Stokey and Lucas (1989). In this paper we study the estimation of structural parameters and their functionals that underlie a class of Markov decision processes (MDP) with discrete controls and time in the infinite horizon setting. Such models are popular in applied work, in particular in labor and industrial organization. The econometrics involved can be seen as an extension of the classical discrete choice analysis to a dynamic framework.

Discrete choice modelling has a long established history in the structural analysis of behavioral economics. McFadden (1974) pioneered the theory and methods of analyzing discrete choice in a static framework. Rust (1987), using *additive separability* and *conditional independence* assumptions, show that a class of dynamic discrete choice models can naturally preserve the familiar structure of discrete choice problems of the static framework. In particular, Rust proposed the Nested Fixed Point (NFP) algorithm to estimate his parametric model by the maximum likelihood method. However, in practice, this method can post a considerable obstacle due to its requirement to repeatedly solve for the fixed point of some nonlinear map to obtain the value functions. The two-step approach of Hotz and Miller (1993) avoided the full solution method by relying on the existence of an inversion map between the normalized value functions and the (conditional) choice probabilities, which significantly reduces the computational burden relative to the NFP algorithm.

The two-step estimator of Hotz and Miller is central to several methodologies that followed, especially in the recent development of the estimation of dynamic games. A class of stationary infinite horizon Markovian games can be defined to include the MDP of interest as a special case. Various estimation procedures have been proposed to estimate the structural parameters of dynamic discrete action games; Pakes, Ostrovsky and Berry (2004), and Aguirregabiria and Mira (2007), considered two-step method of moments and pseudo maximum likelihood estimators respectively, which are included in the general class of asymptotic least square estimators defined by Pesendorfer and Schmidt-Dengler (2008); Bajari, Benkard and Levin (2007) generalizes the simulation-based estimators of Hotz et al. (1994) to the multiple agent setting. However, in both single and multiple agent settings, the aforementioned work assumed the observed state space is finite whenever the transition distribution of the observed state variables is not specified parametrically. As noted by Aguirregabiria and Mira (2002,2007), we should be able to relax this requirement and allow for uncountable observable state space.

In this paper we propose a simple two-step semiparametric approach that falls in the general class of semiparametric estimation discussed in Pakes and Olley (1995), and Chen, Linton and van Keilegom (2003). The criterion function will be based on some conditional moment restrictions that requires consistent estimators of the value functions. The additional difficulty here is due to the fact that the infinite dimensional parameter is defined through a linear integral equation of type II. The study of the statistical properties of solutions to integral equations falls under the growing research area on inverse problem in econometrics, see Carrasco, Florens and Renault (2007) for a survey. Type II integral equations are found, amongst others, in the study of additive models, see Mammen, Linton and Nielson (1995). We show that our problem is generally well-posed and utilize the approach similar to Linton and Mammen (2005) to estimate and provide the distribution theory for the infinite dimensional parameters of interest.

Our estimation strategy can be seen as a generalization of the unifying method of Pesendorfer and Schmidt-Dengler (2008) that allows for continuous components in the observable state space. The novel approach of Pesendorfer and Schmidt-Dengler relies on the attractive feature of the infinite time stationary model, where they write their ex-ante value function as the solution to a matrix equation.[1] We show that the solving of an analogous linear equation, in an infinite dimensional space, is also a well-posed problem for both population and empirical versions (at least for large sample size).[2] We note that an independent working paper of Bajari, Chernozhukov, Hong and Nekipelov (2008) also proposes a sieve estimator for a closely related Markovian games, which allows for continuous observable state space. Therefore our methods are complementary in filling this gap in the literature. However, our estimation strategy, which is simple and intuitive like its predecessor. We use the local approach of kernel smoothing, under some easily interpretable primitive conditions, to provide explicit pointwise distribution theory of the infinite dimensional parameters that would otherwise be elusive with the series or splines expansion. Since the infinite dimensional parameters in MDP are the value functions, they may be of considerable interest themselves. Another advantage for the local estimator includes the optimality in the minimax sense for local linear estimators, see Fan (1993). In addition, we explicitly work under time series framework and provide the type of primitive conditions required for the validity of the methodology.

Since the main idea can be fully illustrated in the single agent setup, for most parts of the paper we consider the single agent setup and leave the discussion of the Markovian game estimation to the latter section. The paper is organized as follows. Section 2 defines the MDP of interest,

[1] A closely related technique is also used in the estimating a dynamic auction game of Jofre-Bonet and Pesendorfer (2003).

[2] We only focus on the estimation aspect as, taking the approach of Magnac and Thesmar (2002), one can simply write down extensions of nonparametric identification results on the per period payoff functions of Pesendorfer and Schmidt-Dengler (2003,2008).

motivates and discusses the estimation strategy and the related linear inverse problem. Section 3 describes in detail the practical implementation of the procedure to obtain the feasible conditional choice probabilities. In Section 4, primitive conditions and the consequent asymptotic distribution are provided, the semiparametric profiled likelihood estimator is illustrated as a special case. Section 5 discusses the extension to dynamic game setting. Section 6 presents a small scale Monte Carlo experiment to study the finite sample performance of our estimator. Section 7 concludes.

2 Markov Decision Processes

We define our time homogeneous MDP and introduce the main model assumptions and notation used throughout the paper. The sources of the computational complexity for estimating MDP are briefly reviewed, there we focus on the representation of the value function as a solution to the policy value equation that can generally be written as an integral equation, in 2.2. We discuss the inverse problem associated with solving such integral equations in 2.3.

2.1 Definitions and Assumptions

We consider a decision process of a forward looking agent who solves the following infinite horizon intertemporal problem. The random variables in the model are the control and state variables, denoted by a_t and s_t respectively. The control variable, a_t, belongs to a finite set of alternatives $A = \{1, \ldots, K\}$. The state variables, s_t, has support $S \subset \mathbb{R}^{L+K}$. At each period t, the agent observes s_t and chooses an action a_t in order to maximize her discounted expected utility. The present period utility is time separable and is represented by $u(a_t, s_t)$. The agent's action in period t affects the uncertain future states according to the (first order) Markovian transition density $p(ds_{t+1}|s_t, a_t)$. The next period utility is subjected to discounting at the rate $\beta \in (0, 1)$. Formally, for any time t, the agent is represented by a triple of primitives (u, β, F), who is assumed to behave according to an optimal decision rule, $\{\alpha_\tau(s_\tau)\}_{\tau=t}^\infty$, in solving the following *sequential problem*

$$V(s_t) = \max_{\{a_\tau(s_\tau)\}_{\tau=t}^\infty} E\left[\sum_{\tau=t}^\infty \beta^\tau u(a_\tau(s_\tau), s_\tau)\middle| s_t\right] \quad \text{s.t.} \quad a_\tau(s_\tau) \in A \text{ for all } \tau \geq t. \tag{1}$$

Under some regularity conditions, see Bertsekas and Shreve (1978) and Rust (1994), Blackwell's Theorem and its generalization ensure the following important properties. First, there exists a stationary (time invariant) Markovian *optimal policy function* $\alpha : S \to A$ so that $\alpha(s_t) = \alpha(s_{t+\tau})$ for any $s_t = s_{t+\tau}$ and any t, τ, where

$$\alpha(s_t) = \arg\max_{a\in A} \{u(a, s_t) + \beta E[V(s_{t+1})|s_t, a_t = a]\}.$$

Secondly, the value function, defined in (1), is the unique solution to the *Bellman's equation*

$$V(s_t) = \max_{a \in A} \{u(a, s_t) + \beta E[V(s_{t+1}) | s_t, a_t = a]\}. \tag{2}$$

We now introduce the following set of modelling assumptions.

ASSUMPTION M1: *(Conditional Independence) The transitional density has the following factorization:* $p(dx_{t+1}, d\varepsilon_{t+1} | x_t, \varepsilon_t, a_t) = q(d\varepsilon_{t+1} | x_{t+1}) f_{X'|X,A}(dx_{t+1} | x_t, a_t)$, *where the first moment of* ε_t *exists and its conditional distribution is absolutely continuous with respect to the Lebesgue measure in* $\mathbb{R}^K$, *we denote its density by* q.

The conditional independence assumption of Rust (1987) is fundamental in the current literature. It is a subject of current research on how to find a practical methodology that can relax this assumption, for example Arcidiacono and Miller (2008). The continuity assumption on the distribution of ε_t ensures we can apply Hotz and Miller's inversion theorem.

ASSUMPTION M2: *The support of* $s_t = (x_t, \varepsilon_t)$ *is* $X \times \mathcal{E}$, *where* X *is a compact subset of* $\mathbb{R}^L$, *in particular,* $x_t = (x_t^c, x_t^d) \in X^C \times X^D$, *and* $\mathcal{E} = \mathbb{R}^K$.

In order to avoid a degenerate model, we assume that the state variables $s_t = (x_t, \varepsilon_t) \in X \times \mathbb{R}^K$ can be separated into two parts, which are observable and unobservable respectively to the econometrician; see Rust (1994a) for various interpretations of the unobserved heterogeneity. Compactness of X is assumed for simplicity, in particular to X^C can be unbounded.

ASSUMPTION M3: *(Additive Separability) The per period payoff function* $u : A \times X \times \mathcal{E} \to \mathbb{R}$ *is additive separable w.r.t. unobservable state variables,* $u(a_t, x_t, \varepsilon_t) = \pi(a_t, x_t) + \sum_{k=1}^{K} \varepsilon_{a_{k,t}} \mathbf{1}[a_t = k]$.

The combination of M1 and M3 allows us to set our model in the familiar framework of static discrete choice modelling.

We shall introduce the structural parameters $\theta \in \Theta \subset \mathbb{R}^M$ that parameterize π later in Section 3 to keep the notation of the general discussion simple. It is indeed our goal to estimate θ as well as some functionals depending on them. Conditions M1 - M3 are crucial to the estimation methodology we propose. These conditions are standard in the literature. In particular, M2 is weaker than the usual finite X assumption when no parametric assumption is assumed on $f_{X'|X,A}(dx_{t+1} | x_t, a_t)$ in the infinite horizon framework. For departures of this framework see the discussion in the survey of Aguirregabiria and Mira (2008) and the references therein. Henceforth Conditions M1 - M3 will be assumed and later strengthened as appropriate.

2.2 Value Functions

Similarly to the static discrete choice models, the choice probabilities play a central role in the analysis of the controlled process. There are two numerical aspects that we need to consider in the evaluation of the choice probabilities. The first are the multiple integrals, that also arise in the static framework, where in practice many researchers avoid this issue via the use of conditional logit assumption of McFadden (1974).[3] The second is regarding the value function - this is unique to the dynamic setup. To see precisely the problem we face, we first update the Bellman's equation (2) under the assumptions M1 - M3,

$$V(s_t) = \max_{a \in A} \{\pi(a, x_t) + \varepsilon_{a,t} + \beta E[V(s_{t+1}) | x_t, a_t = a]\}.$$

Denoting the future expected payoff $E[V(s_{t+1}) | x_t, a_t]$ by $g(a_t, x_t)$, and the choice specific value, net of $\varepsilon_{a,t}$, $\pi(a_t, x_t) + \beta g(a_t, x_t)$ by $v(a_t, x_t)$, the optimal policy function must satisfy

$$\alpha(x_t, \varepsilon_t) = a \Leftrightarrow v(a, x_t) + \varepsilon_{a,t} \geq v(a', x_t) + \varepsilon_{a',t} \text{ for } a' \neq a. \tag{3}$$

The conditional choice probabilities, $\{P(a|x)\}$, are then defined by

$$\begin{aligned} P(a|x) &= \Pr[v(a, x_t) + \varepsilon_{a,t} \geq v(a', x_t) + \varepsilon_{a',t} \text{ for } a' \neq a | x_t = x] \\ &= \int \mathbf{1}[\alpha(x, \varepsilon_t) = a] \, q(d\varepsilon_t | x). \end{aligned} \tag{4}$$

Even if we knew v, (4) will generally not have a closed form and the task of performing multiple integrals numerically can be non-trivial, see Hajivassiliou and Ruud (1994) for an extensive discussion on an alternative approach to approximating integrals. For some specific distributional assumptions on ε_t, for example using the popular i.i.d. extreme value of type I - we can avoid the multiple integrals as (4) has the well known multinomial logit form

$$P(a|x) = \frac{\exp(v(a, x))}{\sum_{a' \in A} \exp(v(a', x))}.$$

Our estimation strategy accommodates for general form of distribution. However, the problem we want to focus on is the fact that we generally do not know v, as it depends on g that is defined through some nonlinear functional equation that we need to solve for. Next, we outline a characterization of the value function that motivates our approach to estimate g (and v).

The main insight to the simplicity of our methodology is motivated from the geometric series representation for the value function that is commonly used in dynamic programming theory, for

[3] Unlike in static models, we do not suffer from the undesirable I.I.A. when use i.i.d. extreme values errors of type I in the dynamic framework.

an example see Bertsekas and Shreve (1978, Chapter 9). More specifically, one can define the value function corresponding to a particular stationary Markovian policy μ by

$$V(s_t;\mu) = E\left[\left.\sum_{\tau=t}^{\infty}\beta^{\tau}u(\mu(s_\tau), s_\tau)\right| s_\tau = s_t\right],$$

which is the solution to the following *policy value equation*

$$V(s_t;\mu) = u(\mu(s_t), s_t) + \beta E[V(s_{t+1};\mu)|s_t].$$

In this paper we only consider values corresponding to the the optimal policy, to reduce the notation, so we suppress the explicit dependence on the policy. Therefore, by definition of the optimal policy, the solution to (2) is also the solution to the following policy value equation

$$V(s_t) = u(\alpha(s_t), s_t) + \beta E[V(s_{t+1})|s_t]. \tag{5}$$

If the state space S is finite, then V is a solution of a matrix equation above since the conditional expectation operator here can be represented by a stochastic transitional matrix. By the dominant diagonal theorem, the matrix representing $(I - \beta E[\cdot|s])_{s\in S}$ is invertible and (5) has a unique solution, solvable by direct matrix inversion or approximated by a geometric series (see the Neumann series below). The notion of simply inverting a matrix has an obvious appeal over Rust's fixed point iterations. In the infinite dimensional case, the matrix equation generalizes to an integral equation. In the presence of some unobserved state variables, we can also define the conditional value function as a solution to the following conditional policy value equation, taking conditional expectation on (5) w.r.t. x_t yields

$$\begin{aligned}
E[V(s_t)|x_t] &= E[u(\alpha(s_t), s_t)|x_t] + \beta E[E[V(s_{t+1})|s_t]|x_t] \\
&= E[u(\alpha(s_t), s_t)|x_t] + \beta E[E[V(s_{t+1})|x_{t+1}]|x_t],
\end{aligned}$$

where the last equality follows from the law of iterated expectations and M1. Noting that, again by M1, $g(a_t, x_t)$ can be written as $E[m(x_{t+1})|x_t, a_t]$, where $m(x_t) = E[V(s_t)|x_t]$, then we have m as a solution to some particular integral equation of type II; more succinctly, m satisfies

$$m = r + \mathcal{L}m, \tag{6}$$

where r is the ex-ante expected immediate payoff given state x_t, namely $E[u(\alpha(s_t), s_t)|x_t = \cdot]$; and the integral operator $\mathcal{L}$ generates discounted expected next period values of its operands, e.g. $\mathcal{L}m(x) = \beta E[m(x_{t+1})|x_t = x]$ for any $x \in X$. If we could solve (6) then we need another level of smoothing on m to obtain the choice specific value v. In particular, we can define g through the following linear transform

$$g = \mathcal{H}m, \tag{7}$$

where $\mathcal{H}$ is an integral operator that generates the choice specific expected next period values of its operands operator, e.g. $\mathcal{H}m(x,a) = \beta E[m(x_{t+1})|x_t = x, a_t = a]$ for any $(x,a) \in X \times A$. Therefore we can write the choice specific value net of unobserved states in a linear functional notation as

$$v = \pi + \beta\mathcal{H}m. \tag{8}$$

In Section 3 we discuss in details on how to use the policy value approach to estimate the model implied transformed of the value functions and choice probabilities.

2.3 Linear Inverse Problems

Before we consider the estimation of v, we need to address some issues regarding the solution of integral equations (6). It is natural to ask the fundamental question whether our problem is well-posed, more specifically, whether the solution of such equation exist and if so, whether it is unique and stable. The study of the solution to such integral equations falls in the general framework of linear inverse problems.

The study of inverse problems is an old problem in applied mathematics. The type of inverse problems one commonly encounters in econometrics are integral equations. Carrasco et al. (2007) focused their discussion on ill-posed problems of integral equations of type I where recent works often needed regularizations in Hilbert Spaces to stabilize their solutions. Here we face an integral equation of type II, which is easier to handle, and in addition, the convenient structure of the policy value equations allows us to easily show that the problem is well-posed in a familiar Banach Space. We now define the normed linear space and the operator of interest, and proof this claim. We shall simply state relevant results from the theory of integral equations. For definitions, proofs and further details on integral equations, readers are referred to Kress (1999) and the references therein.

From the Riesz Theory of operator equations of the second kind with compact operators on a normed space, say $A : X \to X$, we know that $I - A$ is injective if and only if it is surjective, and if it is bijective, then the inverse operator $(I-A)^{-1} : X \to X$ is bounded. For the moment, suppose that X^D is empty, we will be working on the Banach space $(B, \|\cdot\|)$, where $B = C(X)$ is a space of continuous functions defined on the compact subset of $\mathbb{R}^L$, equipped with the sup-norm, i.e. $\|\phi\| = \sup_{x\in X} |\phi(x)|$. $\mathcal{L}$ is a linear map, $\mathcal{L} : C(X) \to C(X)$, such that, for any $\phi \in C(X)$ and $x \in X$,

$$\mathcal{L}\phi(x) = \beta \int_X \phi(x') f_{X'|X}(dx'|x),$$

where $f_{X|X}(dx_{t+1}|x_t)$ denotes the conditional density of x_{t+1} given x_t.

In this case since we know the existence, uniqueness and stability of the solution to (6) are assured for any $r = \phi \in C(X)$ as we can show $\mathcal{L}$ is a contraction. To see this, take any $\phi \in C(X)$ and

$x \in X$,

$$|\mathcal{L}\phi(x)| \leq \beta \int_X |\phi(x')| f_{X'|X}(dx'|x) \leq \beta \sup_{x\in X} |\phi(x)|,$$

since the discounting factor $\beta \in (0,1)$,

$$\|\mathcal{L}\phi\| \leq \beta \|\phi\| \Rightarrow \|\mathcal{L}\| \leq \beta < 1.$$

This implies that our inverse is well-posed. Further, the contraction property means we can represent the solution to (6) using the Neumann series:

$$\begin{aligned} m &= (I - \mathcal{L})^{-1} r \\ &= \lim_{\mathcal{T}\longrightarrow\infty} \sum_{\tau=1}^{\mathcal{T}} \mathcal{L}^{\tau} r. \end{aligned} \tag{9}$$

Therefore the infinite series representation of the inverse suggests one obvious way of approximating the solution to the integral equation which will converge geometrically fast to the true function. If X is countable, then $\mathcal{L}^{\tau}$ would be represented by a τ-step ahead transition matrix (scaled by β^{τ}). Note that the operator for the (uncountable) infinite dimensional case share the analogous interpretation of τ-step ahead transition operator with discounting.

Since our problem is well-posed, then it is reasonable to expect that with sufficiently good estimates of $(r, \mathcal{L}, \mathcal{H})$, our estimated integral equation is also well-posed and will lead to (uniform) consistent estimators for (m, g, v). Our strategy is to use nonparametric methods to generate the empirical versions of (6) and (7), then use them to provide an approximate for v necessary for computing the choice probabilities.

3 Estimation

Given a time series $\{a_t, x_t\}_{t=1}^{T}$ generated from the controlled process of an economic agent represented by (u_{θ_0}, β, p), for some $\theta_0 \in \Theta$, where u_θ reflects the parameterization of π by θ. In this section we provide in details the procedure to estimate θ_0 as well as their corresponding conditional value functions. We based our estimation on the conditional choice probabilities. We define the model implied choice probabilities from a family of value functions, $\{V_\theta\}_{\theta\in\Theta}$, induced by underlying optimal policy that generates the data. In particular, for each θ, V_θ satisfies (cf. equation (5))

$$V_\theta(s_t) = u_\theta(\alpha(s_t), s_t) + \beta E[V_\theta(s_{t+1})|s_t].$$

The policy value V_θ has the interpretation of a discounted expected value for an economic agent whose payoff function is indexed by θ but behaves optimally as if her structural parameter is θ_0. By definition of the optimal policy, V_θ coincides with the solution of a Bellman's equation in (2) when $\theta = \theta_0$. We then define the following (optimal) policy-induced equations to analogous to (6), (7) and (8), respectively for each θ:

$$m_\theta = r_\theta + \mathcal{L} m_\theta, \tag{10}$$

$$g_\theta = \mathcal{H} m_\theta, \tag{11}$$

$$v_\theta = \pi_\theta + \beta \mathcal{H} m_\theta, \tag{12}$$

where r_θ is the ex-ante expected payoff given state x_t, namely $E\left[u_\theta\left(\alpha\left(s_t\right), s_t\right)|x_t = \cdot\right]$; and the integral operators $\mathcal{L}$ and $\mathcal{H}$ are the same as in Section 2.2. The functions m_θ, g_θ and v_θ are defined to satisfy the linear equation and transforms respectively. Naturally, for each $(a, x) \in A \times X$, $P_\theta\left(a|x\right)$ is then defined to satisfy

$$P_\theta\left(a|x\right) = \Pr\left[v_\theta\left(a, x_t\right) + \varepsilon_{a,t} \geq v_\theta\left(a', x_t\right) + \varepsilon_{a',t} \text{ for } a' \neq a | x_t = x\right],$$

which is analogous to (4).

Our methodology proceeds in two steps. In the first step, we nonparametrically compute estimates of the kernels of $\mathcal{L}, \mathcal{H}$ and for each θ, estimate r_θ, which are then used to estimate m_θ by solving the empirical version of the integral equation (10) and estimate g_θ analogously from an empirical version of (11). The second step is the optimization stage, the model implied choice specific value functions are used to compute the choice probabilities that can be used to construct various objective functions to estimate the structural parameter θ_0.

3.1 Estimation of $r_\theta, \mathcal{L}$ and $\mathcal{H}$

There are several decisions to be made to solve the empirical integral equation in (10). We need to first decide on the nonparametric method. We will focus on the method of kernel smoothing due to its simplicity of use as well as its well established theoretical grounding. Our nonparametric estimation of the conditional expectations will be based on the Nadaraya-Watson estimator. However, since we will be working on bounded sets, it is necessary to address the boundary effects. The treatment of the boundary issues is straightforward, the precise trimming condition is described in Section 4. So we will assume to work on a smaller space $X_T \subset X$ where $X_T = \left(X_T^C, X^D\right)$ denotes a set where the support of the uncountable component is some strict compact subset of X^C but increases to X^C in T. When allowing for discrete components we simply use the frequency approach, smoothing over the discrete components is also possible, see the monograph by Li and Racine (2006) for a recent update on this literature. We will also need to make a decision on how to define and interpolate the solution to the empirical version of (10) in practice. We discuss two asymptotically equivalent options for this latter choice, whether the size of the empirical integral equation does or does not depend on the sample size, as one may have a preference given the relative size of the number of observations.

We now define the nonparametric estimators, $(\widehat{r}_\theta, \widehat{\mathcal{L}}, \widehat{\mathcal{H}})$, of $(r_\theta, \mathcal{L}, \mathcal{H})$. Any generic density of a mixed continuous-discrete random vector $w_t = \left(w_t^c, w_t^d\right)$, $f_w : \mathbb{R}^{l^C} \times \mathbb{R}^{l^D} \to \mathbb{R}^+$ for some positive integers l^C and l^D, is estimated as follows,

$$\widehat{f}_w\left(w^c, w^d\right) = \frac{1}{T}\sum_{t=1}^{T} K_h\left(w_t^c - w^c\right)\mathbf{1}\left[w_t^d = w^d\right],$$

where K is some user chosen symmetric probability density function, h is a positive bandwidth and for simplicity independent of w^c. $K_h(\cdot) = K(\cdot/h)/h$ and if $l^C > 1$ then $K_h\left(w_t^c - w^c\right) = \prod_{l=1}^{l^C} K_{h_l}\left(w_{t,l}^c - w_l^c\right)$, $\mathbf{1}[\cdot]$ denotes the indicator function, namely $\mathbf{1}[\mathcal{A}] = 1$ if event $\mathcal{A}$ occurs and takes value zero otherwise. Similar to the product kernel, the contribution from a multivariate discrete variable is represented by products of indicator functions. The conditional densities/probabilities are estimated using the ratio of the joint and marginal densities. The local constant estimator of any generic regression function, $E[z_t|w_t = w]$ is defined by,

$$\widehat{E}[z_t|w_t = w] = \frac{\frac{1}{T}\sum_{t=1}^{T} z_t K_h\left(w_t^c - w^c\right)\mathbf{1}\left[w_t^d = w^d\right]}{\widehat{f}_w(w)}. \tag{13}$$

Estimation of r_θ For any $x \in X_T$,

$$\begin{aligned} r_\theta(x) &= E\left[u_\theta\left(a_t, x_t, \varepsilon_t\right)|x_t = x\right] \\ &= E\left[\pi_\theta\left(a_t, x_t\right)|x_t = x\right] + E\left[\varepsilon_{a_t}|x_t = x\right] \\ &= \rho_{1,\theta}(x) + \rho_2(x). \end{aligned}$$

The first term can be estimated by

$$\widehat{\rho}_{1,\theta}(x) = \sum_{a\in A}\widehat{P}(a|x)\,\pi_\theta(a,x), \tag{14}$$

or, alternatively, the Nadaraya-Watson estimator,

$$\widetilde{\rho}_{1,\theta}(x) = \widehat{E}\left[\pi_\theta\left(a_t, x_t\right)|x_t = x\right].$$

In (14), $\{\widehat{P}(a|x)\}_{a\in A}$ is a kernel estimator of the choice probabilities. We also comment that it might be more convenient to use $\widehat{\rho}_{1,\theta}$ over $\widetilde{\rho}_{1,\theta}$, as we shall see, since the nonparametric estimates for the choice probabilities are required to estimate ρ_2.

The conditional mean of the unobserved states, ρ_2, is generally non-zero due to selectivity. By Hotz and Miller's inversion theorem, we know ρ_2 can be expressed as a known smooth function of the choice probabilities. An estimator of ρ_2 can therefore be obtained by plugging in the local constant (linear) estimator of the choice probabilities. For example, the i.i.d. type I extreme value errors assumption will imply that

$$\rho_2(x) = \gamma + \sum_{a \in A} P(a|x) \log (P(a|x)), \tag{15}$$

where γ is the Euler's constant. Our procedure is not restricted to the conditional logit assumption. Although other distributional assumption will generally not provide a closed form expression for ρ_2 in $\{P(a|x)\}$, it can be computed for any $(a, x) \in A \times X$, for example see Pesendorfer and Schmidt-Dengler (2003) who assume the unobserved states are i.i.d. standard normals. Note also that ρ_2 is independent of θ as the distribution of ε_t is assumed to be known; in principles; our procedure can be written to easily accommodate the case when the conditional distribution of ε_t is known up some finite dimensional parameters.

Estimation of $\mathcal{L}$ and $\mathcal{H}$ For the ease of notation let's suppose X^D is empty. For the integral operators $\mathcal{L}$ and $\mathcal{H}$, if we would like to use the numerical integration to approximate the integral, we only need to provide the nonparametric estimators of their kernels, respectively, $\widehat{f}_{X'|X}(dx_{t+1}|x_t)$ and $\widehat{f}_{X'|X,A}(dx_{t+1}|x_t, a_t)$.

For any $\phi \in C(X_T)$, the empirical operators are defined as,

$$\begin{aligned}
\widehat{\mathcal{L}}\phi(x) &= \int_{X_T} \phi(x') \widehat{f}_{X'|X}(dx'|x), & (16)\\
\widehat{\mathcal{H}}\phi(x,a) &= \int_{X_T} \phi(x') \widehat{f}_{X'|X,A}(dx'|x,a). & (17)
\end{aligned}$$

So $\widehat{\mathcal{L}}$ and $\widehat{\mathcal{H}}$ are linear operators on the Banach space of continuous functions on X_T with range $C(X_T)$ and $C(X_T \times A)$ respectively under sup-norm. Alternatively, we could use the Nadaraya-Watson estimator, defined in (13), to estimate the operators,

$$\begin{aligned}
\widetilde{\mathcal{L}}\phi(x) &= \widehat{E}[\phi(x_{t+1})|x_t = x],\\
\widetilde{\mathcal{H}}\phi(x,a) &= \widehat{E}[\phi(x_{t+1})|x_t = x, a_t = a].
\end{aligned}$$

This approach may be more convenient when sample size is relatively small, and we want to solve the empirical version of (10) by using purely nonparametric methods for interpolation, where we could use the local linear estimator to address the boundary effects.

Note that, if X is finite then the integrals in (16) and (17) will be defined with respect to discrete measures, then $(\widehat{\mathcal{L}}, \widehat{\mathcal{H}})$ and $(\widetilde{\mathcal{L}}, \widetilde{\mathcal{H}})$ can be equivalently represented by the same stochastic matrices.

3.2 Estimation of m_θ, g_θ and v_θ

We first describe the procedure used in Linton and Mammen (2005), by using $(\widehat{\mathcal{L}}, \widehat{\mathcal{H}})$, to solve the empirical integral equation. We define $\widehat{m}_\theta$ as any sequence of random functions defined on X_T that approximately solves $\widehat{m}_\theta = \widehat{r}_\theta + \widehat{\mathcal{L}}\widehat{m}_\theta$. Formally, we shall assume that $\widehat{m}_\theta$ is any random sequence of functions that satisfy

$$\sup_{\theta\in\Theta, x\in X_T} \left| \left(I - \widehat{\mathcal{L}} \right) \widehat{m}_\theta (x) - \widehat{r}_\theta (x) \right| = o_p\left(T^{-1/2}\right), \tag{18}$$

i.e., the right hand side of (18) is approximately zero. We allow this extra generality like Pakes and Pollard (1989) and Linton and Mammen (2005). In practice, we solve the integral equation on a finite grid of points, which reduces it to a large linear system. Next we use $\widehat{m}_\theta$ to define $\widehat{g}_\theta$, specifically we define $\widehat{g}_\theta$ as any random sequence of functions that satisfy

$$\sup_{\theta\in\Theta, a\in A, x\in X_T} \left| \widehat{g}_\theta (a, x) - \widehat{\mathcal{H}}\widehat{m}_\theta (a, x) \right| = o_p\left(T^{-1/2}\right). \tag{19}$$

Once we obtain $\widehat{g}_\theta$, the estimator of v_θ is defined by

$$\sup_{\theta\in\Theta, a\in A, x\in X_T} \left| \widehat{v}_\theta (a, x) - \pi_\theta (a, x) - \beta \widehat{g}_\theta (a, x) \right| = o_p\left(T^{-1/2}\right). \tag{20}$$

For illustrational purposes, ignoring the trimming factors, we will assume that $X = [\underline{x}, \overline{x}] \subset \mathbb{R}$.

For any integrable function ϕ on X, define $J(\phi) = \int \phi(t)\, dt$. Given an ordered sequence of n nodes $\{t_{j,n}\} \subset [a, b]$, and a corresponding sequence of weights $\{\omega_{j,n}\}$ such that $\sum_{j=1}^n \omega_{j,n} = b - a$, a valid integration rule would satisfy

$$\begin{aligned} \lim_{n\to\infty} J_n(\phi) &= J(\phi) \\ J_n(\phi) &= \sum_{j=1}^n \omega_{j,n} \phi(t_{j,n}), \end{aligned}$$

for example Simpson's rule and Gaussian quadrature both satisfy this property for smooth ϕ. Therefore the empirical version of (10) can be approximated for any $x \in [a, b]$ by

$$\widehat{m}_\theta (x) = \widehat{r}_\theta (x) + \beta \sum_{j=1}^n \omega_{j,n} \widehat{f}_{X'|X} (t_{j,n}|x)\, \widehat{m}_\theta (t_{j,n}). \tag{21}$$

So the desired solution that approximately solves the empirical integral equation will satisfy the following equation at each node $\{t_{j,n}\}$,

$$\widehat{m}_\theta (t_{i,n}) = \widehat{r}_\theta (t_{i,n}) + \beta \sum_{j=1}^n \omega_{j,n} \widehat{f}_{X'|X} (t_{j,n}|t_{i,n})\, \widehat{m}_\theta (t_{j,n}).$$

This is equivalent to solving a system of n equations with n variables, the linear system above can be written in a matrix notation as

$$\widehat{\mathbf{m}}_\theta = \widehat{\mathbf{r}}_\theta + \widehat{\mathbf{L}}\widehat{\mathbf{m}}_\theta, \tag{22}$$

where $\widehat{\mathbf{m}}_\theta = (\widehat{m}_\theta(t_{1,n}), \ldots, \widehat{m}_\theta(t_{n,n}))^\top, \widehat{\mathbf{r}}_\theta = (\widehat{r}_\theta(t_{1,n}), \ldots, \widehat{r}_\theta(t_{n,n}))^\top$, I_n is an identity matrix of order n and $\widehat{\mathbf{L}}$ is a square n matrix such that $(\widehat{\mathbf{L}})_{ij} = \beta\omega_{j,n}\widehat{f}_{X'|X}(t_{j,n}|t_{i,n})$. Since $\widehat{f}_{X'|X}(\cdot|x)$ is a proper density for any x, with a sufficiently large n, $(I_n - \widehat{\mathbf{L}})$ is invertible by the dominant diagonal theorem. So there is a unique solution to the system (22) for a given $\widehat{\mathbf{r}}_\theta$. In practice we have a variety of ways to solve for $\widehat{\mathbf{m}}_\theta$ with one obvious candidate being the successive approximation as mentioned in (**??**). Once we obtain $\widehat{\mathbf{m}}_\theta$, we can approximate $\widehat{m}_\theta(x)$ for any $x \in X$ by substituting $\widehat{\mathbf{m}}_\theta$ into the RHS of (21). This is known as the Nyström interpolation. We need to approximate another integral to estimate g_θ. This could be done using the conventional method of kernel regression as discussed in Section 3.1, or by appropriately selecting sequences of r nodes $\{q_{j,r}\}$ and weights $\{\zeta_{j,n}\}$ so that

$$\widehat{g}_\theta(j, x) = \sum_{j=1}^{r} \zeta_{j,n}\widehat{f}_{X'|X,A}(q_{j,r}|x, j)\,\widehat{m}_\theta(q_{j,r}),$$

where the computation for this last linear transform is trivial. See Judd (1998) for a more extensive review of the methods and issues of approximating integrals and also the discussion of iterative approaches in Linton and Mammen (2003) for large grid sizes.

Alternatively, we can form a matrix equation of size $T-1$,

$$\widetilde{\mathbf{m}}_\theta = \widetilde{\mathbf{r}}_\theta + \widetilde{\mathbf{L}}\widetilde{\mathbf{m}}_\theta,$$

to estimate equation (10) at the observed points with the t-th element. For each t, let

$$\widetilde{m}_\theta(x_t) = \widetilde{r}_\theta(x_t) + \beta\frac{\frac{1}{T-1}\sum_{t=1}^{T-1}\widetilde{m}_\theta(x_{t+1})K_h(x_t - x)}{\frac{1}{T-1}\sum_{\tau=1}^{T-1}K_h(x_\tau - x_t)}.$$

By the dominant diagonal theorem, the matrix equation above always has a unique solution for any $T \geq 2$. Once solved, the estimators of $\widetilde{m}_\theta$ can be interpolated by

$$\widetilde{m}_\theta(x) = \widetilde{r}_\theta(x) + \beta\widehat{E}[m(x_{t+1})|x_t = x],$$

for any $x \in X_T$. Similarly, $\widetilde{g}_\theta$ and $\widetilde{v}_\theta$ can be estimated nonparametrically without introducing any additional numerical error. Clearly, the more observation we have, the latter method will be more difficult as dimension of the matrix representing $\widetilde{\mathbf{L}}$ is large whilst the grid points for the former empirical equation is user-chosen.

3.3 Estimation of θ

By construction, when $\theta = \theta_0$, the model implied conditional choice probability P_θ coincides with the underlying choice probabilities defined in (4). Therefore one natural estimator for the finite dimensional structural parameters can be obtained by maximizing a likelihood criterion. Define

$$Q_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log P_\theta(a_t|x_t); \quad \widehat{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \mathfrak{c}_{t,T} \log \widehat{P}_\theta(a_t|x_t). \tag{23}$$

Here $\{\mathfrak{c}_{t,T}\}$ is a triangular array of trimming factors, more discussion on this can be found in Section 4. In practice, we replace $P_\theta(a|x)$ by

$$\widehat{P}_\theta(a|x) = \Pr\left[\widehat{v}_\theta(a, x_t) + \varepsilon_{a,t} \geq \widehat{v}_\theta(a', x_t) + \varepsilon_{a',t} \text{ for } a' \neq a | x_t = x\right],$$

where $\widehat{v}_\theta$ satisfies condition (20). Of particular interest is the special case of the conditional logit framework, as discussed in Section 2, where we have

$$\widehat{P}_\theta(a|x) = \frac{\exp(\widehat{v}_\theta(a, x))}{\sum_{a' \in A} \exp(\widehat{v}_\theta(a', x))}.$$

Therefore $\widehat{Q}_T$ denotes the feasible objective function, which is identical to Q_T when the infinite dimensional component $\widehat{v}_\theta$ is replaced by v_θ. We define our maximum likelihood estimator, $\widehat{\theta}$, to be any sequence that satisfy the following in equality

$$\widehat{Q}_T(\widehat{\theta}) \geq \sup_{\theta \in \Theta} \widehat{Q}_T(\theta) - o_p\left(T^{-1/2}\right). \tag{24}$$

Alternatively, a class of criterion functions can be generated from the following conditional moment restrictions

$$E\left[\mathbf{1}[a_t = a] - P_\theta(a|x_t) | x_t\right] = 0 \text{ for all } a \in A \text{ when } \theta = \theta_0.$$

Note that these moment conditions are the infinite dimensional counterparts (with respect to the observable states) of equation (18) in Pesendorfer and Schmidt-Dengler (2008) for a single agent problem.

There are general large sample theory of profiled semiparametric estimators available that treat the estimators defined in our models.[4] In particular, the work of Pakes and Olley (1995) and Chen, Linton and van Keilegom (2003) provide high level conditions for obtaining root$-T$ consistent estimators are directly applicable. The latter is a generalization of the work by Pakes and Pollard (1989), who provided the asymptotic theory when the criterion function is allowed to be non-smooth, which

[4] We can generally write the objective functions from the likelihood criterion (via first order conditions) and the

may arise if we use simulation methods to compute the multiple integral of (4), to the semiparametric framework.

In Section 4, as an illustration, we derive the asymptotic distribution of the semiparametric likelihood estimator under a set of weak conditions in the conditional logit framework.

3.4 Practical Discussion

We reflect on the computational effort required of the proposed method. We only discuss the estimation of the conditional value functions. It will be helpful to have in mind the methodology of Pesendorfer and Schmidt-Dengler (2008) as our methods coincide when the X is finite and there is only 1 player in the game (vice versa, extending from a single agent decision process to a dynamic game). For each θ, the nonparametric estimates of $(r_\theta, \mathcal{L}, \mathcal{H})$ have closed form and are very easy to compute even with large dimensions, further, the empirical integral operators (or their approximations) only need to be computed once at the required nodes since they do not depend on θ. Solving the empirical integral equation to obtain $\widehat{m}_\theta$, in (22), is the only potential complication that does not exist in a static problem. However, in this setup, this reduces to the need to invert a large matrix that approximates $(I - \mathcal{L})$ that only need to be done once at the beginning and stored for future computation with any other θ. Estimators of $(m_\theta, g_\theta, v_\theta)$ are obtained trivially for any θ, by simple matrix multiplication, once the empirical operator of $(I - \mathcal{L})^{-1}$ is obtained. We note that further computational gain is possible if π_θ is linear in θ. More specifically, if $\pi_\theta = \theta^\top \pi_0$ for some known functions π_0 then $r_\theta = \theta^\top r_0 + \rho_2$, where $r_0(\cdot) = \sum_{a \in A} P(a|\cdot)\, \pi_0(a, \cdot)$. Utilizing the fact that the inverse of $(I - \mathcal{L})$ is a linear operator, we have $m_\theta = \theta^\top (I - \mathcal{L})^{-1} r_0 + (I - \mathcal{L})^{-1} \rho_2$, where the estimates of $(I - \mathcal{L})^{-1} r_0$ and $(I - \mathcal{L})^{-1} \rho_2$ only need to be computed once. See Hotz, Miller, Sanders and Smith (1994) and Bajari, Benkard and Levin (2007) for related utilization of the repeated substitution concept.

However, it is important to note that, as we have decided on the kernel smoothing approach there is an issue of bandwidth selection which is important for small sample properties. Further, it is easy

moment restrictions in the following way:

$$M_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} q(a_t, x_t; \theta, v_\theta); \quad \widehat{M}_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \mathfrak{c}_{t,T}\, q(a_t, x_t; \theta, \widehat{v}_\theta),$$

where $\widehat{M}_T$ is the feasible counterpart of M_T. Define the limiting objective function $M(\theta) = \lim_{T \to \infty} E M_T(\theta)$, which is assumed to exist and is uniquely minimized at $\theta = \theta_0$. We then define our estimator to be any sequence that satisfy the following inequality,

$$\left\| \widehat{M}_T(\widehat{\theta}) \right\| \leq \inf_{\theta \in \Theta} \left\| \widehat{M}_T(\theta) \right\| + o_p\left(T^{-1/2}\right).$$

to see that the invertibility of the matrix $(I - \widehat{\mathbf{L}})$ and $(I - \widetilde{\mathbf{L}})$ are not dependent on the number of continuous and/or discrete components. Clearly, there are a lot of choices available regarding integral approximation and matrix inversion methods. It is beyond the scope of this paper to analyze the finite sample performance of these various methodologies.

4 Distribution Theory

In this section we provide a set of primitive conditions and derive the distribution theory for the estimators $\widehat{\theta}$, as defined in (24), and $(\widehat{m}_\theta, \widehat{g}_\theta)$ as defined in (18) and (19) respectively when the unobserved state variables is distributed as i.i.d. extreme value of type I. This distributional assumption is the most commonly used in practice as it yields closed-form expressions for the choice probabilities. We also restrict the dimensionality of X^C to be a subset of $\mathbb{R}$, the reason being this is the scenario that applied researchers may prefer to work with. These specifics do not limit the usefulness of the primitives provided. For other estimation criteria, since two-step estimation problems of this type can be compartmentalized into nonparametric first stage and optimization in the second stage, the primitives below will be directly applicable. In particular, the discussions and results in 4.1 are independent of the choice of the objective functions chosen in the second stage. There might be other intrinsically continuous observable state variables that require discretizing but with increasing dimension in X^C, the practitioners will need to employ higher order kernels and/or undersmooth in order to obtain the parametric rate of convergence for the finite structural parameters, adaptation of the primitives are straightforward and will be discussed accordingly.

4.1 Infinite Dimensional Parameters

The relevant large sample properties for the nonparametric first stage, under the time series framework, for the pointwise results see the results of Roussas (1967,1969), Rosenblatt (1970,1971) and Robinson (1983). Roussas first provided central limit results for kernel estimates of Markov sequences, Rosenblatt established the asymptotic independence and Robinson generalized such results to the α-mixing case. The uniform rates have been obtained for the class of polynomial estimators by Masry (1996), in particular, our method is closely related to the recent framework of Linton and Mammen (2005) who obtained the uniform rates and pointwise distribution theory for the solution of a linear integral equation of type II.

We begin with some primitives. In addition to M1 - M3, they are not necessary and only sufficient but they are weak enough to accommodate most of the existing empirical works in applied labor and industrial organization involving estimation of MDP.

We denote the strong mixing coefficient as

$$\alpha(k) = \sup_{t \in \mathbb{N}} \sup_{A \in \mathcal{F}_{t+k}^{\infty}, \mathcal{F}_{-\infty}^{t}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \quad \text{for } k \in \mathbb{Z},$$

where $\mathcal{F}_a^b$ denotes the sigma-algebra generated by $\{a_t, x_t\}_{t=a}^b$. Our regularity conditions are listed below:

B1 *$X \times \Theta$ is a compact subset of $\mathbb{R}^J \times \mathbb{R}^L$ with $X^C = [\underline{x}, \overline{x}]$.*

B2 *The process $\{a_t, x_t\}_{t=1}^T$ is strictly stationary and strongly mixing, with a mixing coefficient $\alpha(k)$,such that for some $C \geq 0$ and some, possibly large $\chi > 0, \alpha(k) \leq Ck^{-\chi}$.*

B3 *The density of x_t is absolutely continuous $f_{X^C,X^D}(dx_t, x_t^d)$ for each $x_t^d \in X^D$. The joint density of (a_t, x_t) is bounded away from zero on X^C and is twice continuously differentiable over X^C for each $(x_t^d, a_t) \in X^D \times A$. The joint density of (x_{t+1}, x_t, a_t) is twice continuously differentiable over $X^C \times X^C$ for each $(x_{t+1}^d, x_t^d, a_t) \in X^D \times X^D \times A$.*

B4 *The mean of the per period payoff function $u_\theta(a_t, x_t)$ is twice continuously differentiable on $X^C \times \Theta$ for each $(x_t^d, a_t) \in X^D \times A$.*

B5 *The kernel function is a symmetric probability density function with bounded support such that for some constant $C, |K(u) - K(v)| \leq C|u - v|$. Define $\mu_j(K) = \int u^j K(u)\,du$ and $\kappa_j(K) = \int K^j(u)\,du$.*

B6 *The bandwidth sequence h_T satisfies $h_T = \gamma_0(T)T^{-1/5}$ and $\gamma_0(T)$ bounded away from zero and infinity.*

B7 *The triangular array of trimming factors $\{\mathfrak{c}_{t,T}\}$ is defined such that $\mathfrak{c}_{t,T} = \mathbf{1}[x_t^c \in X_T^C]$ where $X_T = [\underline{x} + c_T, \overline{x} - c_T]$ and $\{c_T\}$ is any positive sequence converging monotonically to zero such that $h_T < c_T$.*

B8 *The distribution of ε_t is known to be distributed as i.i.d. extreme value of type I across K alternatives, and is mean independent of x_t and is i.i.d. across t.*

The compactness of the parameter space in B1 is standard. Compactness of the continuous component of the observable state space can be relaxed by using an increasing sequence of compact sets that cover the whole real line, see Linton and Mammen (2005) for the modelling in the tails of the distribution. The dimension of X^C is assumed to be 1 for expositional simplicity, discussion on this is follows the theorems below. On the other hand, it is a trivial matter to add arbitrary (finite) number of discrete components to X^D.

Condition B2 is quite weak despite the value of χ can be large.

The assumptions of B3, B4 and B5 are standard in the kernel smoothing literature using second order kernel.

Here in B6 we use the bandwidth with the optimal MSE rate for a regular 1-dimensional nonparametric estimates.

The trimming factor in B7 provides the necessary treatment of the boundary effects. This would ensure all the uniform convergence results on the expanding compact subset $\{X_T\}$ whose limit is X. In practice we will want to minimize the trimming out of the data, we can choose c_T close enough to h_T to do this.

Condition B8 is not necessary for consistency and asymptotic normality for any of the parameters below. The only requirement on the distribution of ε_t, for our methodology to work, is that it allows us to employ Hotz and Miller's inversion theorem. A sufficient condition for that is the distribution of ε_t is known and satisfy M1. In particular, B8 yields us the simple multinomial logit form that is often used in practice. For other distribution will result in the use of a more complicated inversion map, for example see Pesendorfer and Schmidt-Dengler (2003) for the Gaussian case.

Next we provide pointwise distribution theory for the nonparametric estimators obtained from the first stage, as described in Section 3, for any given set of values of the structural parameters. The bias and the variance terms are complicated, the explicit formulae can be found along with all proofs in the Appendix.

THEOREM 1. *Suppose $B1-B8$ hold. Then for each $\theta \in \Theta$, there exists deterministic functions $\eta_{m,\theta}$ and $\omega_{m,\theta}$ such that for each $x \in \operatorname{int}(X)$,*

$$\sqrt{Th_T}\left(\widehat{m}_\theta(x) - m_\theta(x) - \frac{1}{2}\mu_2 h_T^2 \eta_{m,\theta}(x)\right) \Longrightarrow \mathcal{N}\left(0, \omega_{m,\theta}(x)\right),$$

where $\widehat{m}_\theta(x)$ is defined as in (18) and:

$$\begin{aligned} \eta_{m,\theta}(x) &= (I-\mathcal{L})^{-1}\left(\eta_{r,\theta} + \eta_{\mathcal{L},\theta}\right)(x), \\ \omega_{m,\theta}(x) &= \frac{\kappa_2}{f_X(x)}\left(\beta^2 \operatorname{var}\left(m_\theta(x_{t+1})\,|x_t = x\right) + \omega_{r,\theta}(x)\right). \end{aligned}$$

Some components of the bias and variance are complicated, in particular the explicit form of $\eta_{r,\theta}$, $\eta_{\mathcal{L},\theta}$ and $\omega_{r,\theta}$ can be found below in (39),(48) and (40) respectively. The estimators $\widehat{m}_\theta(x)$ and $\widehat{m}_\theta(x')$ are also asymptotically independent for any $x \neq x'$. Furthermore,

$$\sup_{(x,\theta)\in X_T\times\Theta} \left|\widehat{m}_\theta(x) - m_\theta(x)\right| = o_p\left(T^{-1/4}\right).$$

The pointwise rate of convergence, $T^{-2/5}$, is the usual optimal rate (in the MSE sense) of a $1-$dimensional nonparametric function. The above is obtained by using analogous arguments of

Linton and Mammen (2005) after showing that the conditional density estimator that define the empirical integral operator converges uniformly (see Masry (1996)) over its domain. Similar to Theorem 1, we also obtain the following results for the estimator of g_θ.

THEOREM 2. *Suppose $B1-B8$ hold. Then for each $\theta\in\Theta, x\in\operatorname{int}(X)$ and $a\in A$,*

$$\sqrt{Th_T}\left(\widehat{g}_\theta(a,x)-g_\theta(a,x)-\frac{1}{2}\mu_2h_T^2\eta_{g,\theta}(a,x)\right)\Longrightarrow\mathcal{N}\left(0,\omega_{g,\theta}(a,x)\right),$$

where $\widehat{g}_\theta(a,x)$ is defined as in (19) and:

$$\begin{aligned}\eta_{g,\theta}(a,x) &= \mathcal{H}(I-\mathcal{L})^{-1}\left(\eta_{r,\theta}+\eta_{\mathcal{L},\theta}\right)(a,x)+\eta_{\mathcal{H},\theta}(a,x),\\ \omega_{g,\theta}(a,x) &= \frac{\kappa_2}{f_{X,A}(x,a)}\operatorname{var}\left(m_\theta(x_{t+1})\,|x_t=x,a_t=a\right).\end{aligned}$$

The explicit form of $\eta_{r,\theta}$, $\eta_{\mathcal{L},\theta}$ and $\eta_{\mathcal{H},\theta}$ can be found in (39),(48) and (49) respectively. $\widehat{g}_\theta(a,x)$ and $\widehat{g}_\theta(a',x')$ are also asymptotically independent for any $x\neq x'$ and any a. Furthermore,

$$\sup_{(x,a,\theta)\in X_T\times A\times\Theta}\left|\widehat{g}_\theta(a,x)-g_\theta(a,x)\right|=o_p\left(T^{-1/4}\right).$$

We end with a brief discussion of the change in primitives required to accommodate the case when the dimension of X^C is higher than 1. Clearly, using the optimal (MSE) rates for h_T, $\dim\left(X^C\right)$ cannot exceed 3 with second order kernel if we were to have the uniform rate of convergence for our nonparametric estimates to be faster than $T^{-1/4}$ that is necessary for $\sqrt{T}-$consistency of the finite dimensional parameters. It is possible to overcome this by exploiting additional smoothness (if available) of our densities. This can be done by using higher order kernels to control the order of the bias, for details of their constructions and usages see Robinson (1988) and also Powell, Stock and Stoker (1989).

4.2 Finite Dimensional Parameters

In order to obtain consistency result and the parametric rate of convergence for $\widehat{\theta}$, we need to adjust some assumptions described in the previous subsection and add an identification assumption. Consider:

B6′ *The bandwidth sequence h_T satisfies $Th_T^4\to 0$ and $Th_T^2\to\infty$*

B9 *The value $\theta_0\in\operatorname{int}(\Theta)$ is defined by, for any $\varepsilon>0$*

$$\sup_{\|\theta-\theta_0\|\geq\varepsilon}Q(\theta_0)-Q(\theta)>0,$$

where $Q(\theta)$ denotes the limiting objective function of Q_T (defined in (23)), namely $Q(\theta)=\lim_{T\to\infty}EQ_T(\theta)$.

The rate of undersmoothing (relative to B6) in Condition B6$'$ ensures that the bias from the nonparametric estimation disappears sufficiently quickly to obtain parametric rate of convergence for $\widehat{\theta}$. To accommodate for higher dimension of X^C, we generally cannot just proceed by undersmoothing but combining this with the use higher order kernels, again, see Robinson (1988) and also Powell, Stock and Stoker (1989).

Condition B9 assumes the identification of the parametric part. This is a high level assumption that might not be easy to verify due to the complication with the value function. In practice we will have to check for local maxima for robustness. We note that this is the only assumption concerning the criterion function, for other type of objective functions, obvious analogous identification conditions will be required.

The properties of $\widehat{\theta}$ can be obtained by application of the asymptotic theory for semiparametric profile estimators. This requires uniform expansion $\widehat{g}_\theta$ (and hence $\widehat{m}_\theta$) and their derivatives with respect to θ.

Theorem 3. *Suppose* $B1-B5, B6'$ and $B7-B9$ *hold. Then*

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \Longrightarrow \mathcal{N}\left(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right),$$

where $\mathcal{I}$ *is a complicated term representing the asymptotic variance of the leading terms in* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q\left(a_t,x_t;\theta_0,\widehat{g}_{\theta_0}\right)}{\partial\theta}$ *(see Appendix A) and*

$$\mathcal{J}=E\left[\frac{\partial^2 q\left(a_t,x_t;\theta_0,g_{\theta_0}\right)}{\partial\theta\partial\theta^\top}\right].$$

The root-T rate of convergence is common for such semiparametric estimators when the dimension of the continuous component of X is not too large under some smoothness assumptions. We next present the results for the feasible estimators of m and g.

Theorem 4. *Suppose* $B1-B5, B6'$ *and* $B7-B9$ *hold. Then for any arbitrary estimator* $\widetilde{\theta}$ *such that* $||\widetilde{\theta}-\theta_0||=O_p\left(T^{-1/2}\right)$ *and* $x\in\operatorname{int}(X)$,

$$\sqrt{Th_T}\left(\widehat{m}_{\widetilde{\theta}}(x)-m_{\theta_0}(x)\right)\Longrightarrow\mathcal{N}\left(0,\omega_{m,\theta_0}(x)\right),$$

where $\widehat{m}_\theta, \eta_{m,\theta}$ *and* $\omega_{m,\theta}$ *are defined as those in Theorem 1 and,* $\widehat{m}_{\widetilde{\theta}}(x)$ *and* $\widehat{m}_{\widetilde{\theta}}(x')$ *are asymptotically independent for any* $x\neq x'$.

Similarly, for g_θ we have the following result.

Theorem 5. *Suppose* $B1-B5, B6'$ *and* $B7-B9$ *hold. Then for any arbitrary estimator* $\widetilde{\theta}$ *such that* $||\widetilde{\theta}-\theta_0||=O_p\left(T^{-1/2}\right), x\in\operatorname{int}(X)$ *and* $a\in A$,

$$\sqrt{Th_T}\left(\widehat{g}_{\widetilde{\theta}}(a,x)-g_{\theta_0}(a,x)\right)\Longrightarrow\mathcal{N}\left(0,\omega_{g,\theta_0}(a,x)\right),$$

where $\widehat{g}_\theta, \eta_{g,\theta}$ and $\omega_{g,\theta}$ are defined as those in Theorem 2 and, $\widehat{g}_{\widetilde{\theta}}(a,x)$ and $\widehat{g}_{\widetilde{\theta}}(a',x')$ are asymptotically independent for any $x \neq x'$ and any a.

Given the explicit forms of the bias and variance terms provided in the above theorems, inference can be conducted using large sample approximation based on obvious plug-in estimators. However, due to their complicated form, bootstrap procedures would most likely be preferred in practice. Nevertheless, the explicit expressions we derive in the Appendix are still useful as they provide insights into to the variation in the bias and variance of our estimators.

5 Markovian Games

The development of empirical dynamic games is of recent interest especially in the industrial organization literature. See Ackerberg et al. (2005) for an excellent survey. In this section we briefly summarize how we can use the methodology discussed in previous sections to estimate a class of Markovian games. Similar to Bajari et al. (2008), we consider the same class of dynamic games described in Aguirregabiria and Mira (2007), Bajari et al. (2007), Pakes et al. (2004) and Pesendorfer and Schmidt-Dengler (2008), and allowing the observable state variable to have continuous component. We refer the reader to our working paper version, Srisuma and Linton (2009), for a detailed discussion.

First, note that Pesendorfer and Schmidt-Dengler's (2008) results on the characterization (Proposition 1) and the existence (Theorem 1) of the Markov perfect equilibrium can be readily extended to this more general framework.[5] To avoid repetition, we proceed directly to the policy value equation for each player i, induced by the equilibrium best responses, $\{\alpha_i\}_{i=1}^N$, which generate the observed data. For any θ, we have

$$V_{i,\theta}(s_{it}) = E\left[u_{i,\theta}(\alpha_i(s_{it}), a_{-it}, s_{it})\,|s_{it}\right] + \beta E\left[V_{i,\theta}(s_{it+1})\,|s_{it}\right],$$

where a_{-it} denotes the usual notation of the actions of all other players except player i. Following the arguments in Section 3, where x_t now represents observed public information (to all the players and econometricians), the policy value equation can be used to obtain its conditional counterpart

$$E\left[V_{i,\theta}(s_{it})\,|x_t\right] = E\left[u_{i,\theta}(\alpha_i(s_{it}), a_{-it}, s_{it})\,|x_t\right] + \beta E\left[E\left[V_{i,\theta}(s_{it+1})\,|x_{t+1}\right]|x_t\right],$$

[5] This follows since: (Proposition 1) the arguments in the proof of their Proposition 1 is done pointwise on the support of the observable state space; (Theorem 1) Their equation (11) (on page 909) becomes a continuous functional equation in an infinite dimensional space. We can appeal to fixed point theorems in infinite dimensional spaces under some weak smoothness conditions on the primitve functions (such as Schauder or Tikhonov fixed point theorems, see Granas and Dugundji (2003)).

where we utilize an analogue of the conditional independence assumption, in M1, for the multi-agent case. The model implied choice specific value function for each player, denoted by $v_{i,\theta}$ (see below), can then again be written as a linear transform of the solution to an integral equation (the conditional policy value equation), namely

$$E\left[V_{i,\theta}\left(s_{it+1}\right)|x_t, a_{it}\right] = E\left[E\left[V_{i,\theta}\left(s_{it+1}\right)|x_{t+1}\right]|x_t, a_{it}\right],$$

which can then be used to define the model implied choice probabilities. More specifically, we define

$$P_{i,\theta}\left(a|x_t\right) = \int \mathbf{1}\left[\alpha_{i,\theta}\left(s_{it}\right) = a\right] q_i\left(d\varepsilon_{it}|x_t\right),$$

where ε_{it} denotes a vector of player i's private information, and under additive separability assumption

$$\begin{aligned} \alpha_{i,\theta}\left(s_{it}\right) &= a \Leftrightarrow v_{i,\theta}\left(a, x_t\right) + \varepsilon_{a,it} \geq v_{i,\theta}\left(a', x_t\right) + \varepsilon_{a',it} \text{ for all } a' \in A, \\ v_{i,\theta}\left(a, x\right) &= E\left[\pi_{i,\theta}\left(a, a_{-it}, x_t\right)|x_t = x, a_{it} = a\right] + \beta_i E\left[V_{i,\theta}\left(s_{it+1}\right)|x_t = x, a_{it} = a\right]. \end{aligned}$$

By assuming that the data is generated from a single Markov perfect equilibrium, the conditional expectations (on the observable) are defined with respect to some particular equilibrium distributions that are nonparametrically identified. We can define two-step estimators for the finite dimensional structural parameters from the estimates of $\{P_{i,\theta}\}_{i=1}^N$, by feasible likelihood criterion function or other minimum distance objective functions derived from

$$E\left[\mathbf{1}\left[a_{it} = a\right] - P_{i,\theta}\left(a|x_t\right)|x_t\right] = 0 \text{ for all } a \in A \text{ and } i \in \{1, \ldots, N\} \text{ when } \theta = \theta_0,$$

which are, precisely, the infinite dimensional counterparts of the matrix equation (18) in Pesendorfer and Schmidt-Dengler (2008).

In terms of the practical implementation of estimating the nonparametric functions in the first stage, note that we can write the conditional policy value function in a familiar way

$$m_{i,\theta} = r_{i,\theta} + \mathcal{L} m_{i,\theta},$$

where $(r_{i,\theta}, \mathcal{L})$ are $E\left[u_{i,\theta}\left(\alpha_i\left(s_{it}\right), a_{-it}, s_{it}\right)|x_t\right]$ and the conditional expectation (integral) operator representing the transition of the public information. By the same arguments used in Section 2.3, $\mathcal{L}$ is a contraction map. Even if each player has distinct discounting factor, $\{\mathcal{L}_i\}_{i=1}^N$, these integral equations can be estimated and solved independently. Therefore the comments and discussions in Section 3.4 directly apply to the game setting as well.

6 Numerical Illustration

In this section we illustrate some finite sample properties of our proposed estimator in a small scale Monte Carlo experiment.

Design and Implementation:

We consider the decision process of an agent (say, a mobile store vender) who, in each period t, has a choice to operate in either location A or B. The decision variable a_t takes value 1 if location A is chosen, and 0 otherwise. The immediate payoff from the decision is

$$u\left(a_t, x_t, \varepsilon_t\right) = \pi_{\theta_0}\left(a_t, x_t\right) + a_t \varepsilon_{1,t} + \left(1 - a_t\right) \varepsilon_{2,t},$$

where $\pi_\theta\left(a_t, x_t\right) = \theta_1 a_t x_t + \theta_2\left(1 - a_t\right)\left(1 - x_t\right)$. Here x_t denotes a publicly observed measure of the demand determinant that has been normalized to lie between $[0,1]$. The vector $\left(\varepsilon_{1,t}, \varepsilon_{2,t}\right)$ represents some non-persistant idiosyncratic private costs associated with each choice, which are distributed as i.i.d. extreme value of type 1, that are not observed by the econometricians. To capture the most general aspect of the decision processes discussed in the paper, the future value of x_{t+1} evolves stochastically and its conditional distribution is affected by the observables from the previous period $\left(a_t, x_t\right)$. We suppose the transition density has the following form

$$f_{X'|X,A}\left(x'|x, a\right) = \begin{cases} \delta_{11}\left(x\right) x' + \delta_{12}\left(x\right) & \text{when } a = 1 \\ \delta_{21}\left(x\right) x' + \delta_{22}\left(x\right) & \text{when } a = 0 \end{cases}.$$

We design our model to be consistent with a plausible scenario that future demand builds on existing demand, which is driven by whether or not the vender was present at a particular location. In particular, if the demand at location A is high and the vender is not present, the demand at location A is more likely to be significantly lower for the next period (and vice versa). We use the following simple specific forms for $\{\delta_{ij}\}$ that display such behavior

$$\begin{array}{ll} \delta_{11}\left(x\right) = 2\left(2x - 1\right), & \delta_{12}\left(x\right) = 2\left(1 - x\right) \\ \delta_{21}\left(x\right) = 2\left(1 - 2x\right), & \delta_{22}\left(x\right) = 2x \end{array}.$$

To introduce some asymmetry, we impose that the agent has underlying preference towards location A, which is captured by the condition $\theta_{01} > \theta_{02} > 0$.

We set $\left(\beta, \theta_{01}, \theta_{02}\right)$ to be $(0.9, 1, 0.5)$ and use the fixed point method described in Rust (1996) to generate the controlled Markovian process under the proposed primitives. The initial state is taken as $x_1 = 1/2$, we begin sampling each decision process after 1000 periods and consider $T = 100, 500, 1000$. We conduct 1000 replications of each time length. For each T, we obtain our estimators by following the procedure described in Section 3. To approximate the integral equation we partition $[0,1]$ by using 1000 equally-spaced grid points. Since the support of the observable state is compact, we need to trim of values near the boundary. As an alternative, we employ a simple boundary corrected kernel, see Wand and Jones (1994), based on a Gaussian kernel, namely

$$K_h^b\left(x_t - x\right) = \begin{cases} \frac{1}{h} K\left(\frac{x_t - x}{h}\right) / \int_{-x/h}^{\infty} K\left(v\right) dv & \text{if} \quad x \in [0, h) \\ \frac{1}{h} K\left(\frac{x_t - x}{h}\right) & \text{if} \quad x \in (h, 1-h) \\ \frac{1}{h} K\left(\frac{x_t - x}{h}\right) \int_{-\infty}^{(1-x)/h} K\left(v\right) dv & \text{if} \quad x \in (1-h, 1] \end{cases},$$

where K is the pdf of a standard normal. We consider three choices of bandwidth $\left(h, h^{1/2}, h^{3/2}\right)$, where $h = 1.06 s T^{-1/5}$ and s denotes the standard deviation of the observed $\{x_t\}$.

Results:

We report the summary statistics for the $\widehat{\theta}_1$ and $\widehat{\theta}_2$ in Table 1 and 2 respectively. The bias (mean and median) and the standard deviation all decrease as the sample size increases across all bandwidths. The ratio of the bias and standard deviation remain small even at larger sample size, providing support that our estimators are consistent. The ratio of the scaled interquatile range (by a factor 1.349) and the standard error are also close to 1, which is a trait of a normal random variable.

We also provide the estimated mean deviations of $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t\right]$ for various values of x_t when $a_t = 1$ and 2 in Table 3 and 4 respectively. We find that the bias near the boundaries are larger relative to the interior and the interquartile range of the estimates decrease with sample size across all bandwidths. Generally the absolute values of the bias are small and decreasing as sample size increases. However, the biases corresponding to the bandwidth $h^{3/2}$ appear to not necessarily decrease with sample sizes, especially closer to the boundaries. This could be due to a particular discretization effects.

In regards to the relations between the point and function estimators, since the conditional choice probabilities only depend on the difference between choice specific values, we consider the estimated differences in Figures 1 - 3. The figures contain the graphs of the true and its mean differences plotted on the same scale, for different bandwidths, across different sample sizes. The plots show that the bias is larger near the boundary, and they decrease as sample size increases for all bandwidths.

7 Conclusion

In this paper, we provide a method to estimate a class of Markov decision processes that allows for continuous observable state space. The type of primitive conditions are provided for the inference of the finite and infinite dimensional parameters in the model. Our estimation technique relies on the convenient well-posed linear inverse problem presented by the policy value equation. It inherits the computational simplicity of Pesendorfer and Schmidt-Dengler (2008) that is independent of the parameterization of the per period utility function. We also illustrate how this method can be extended naturally to the estimation of Markovian games in a similar setting to that of Bajari et al. (2008) Their identification results directly apply here.

There are some practical aspects of our estimators worth exploring. Firstly, is the role of numerical error brought upon by approximating the integral in the case that we have large sample size compared to the purely nonparametric approximation. Second is to see how our estimator performs in practice relative to discretization methods. Thirdly, some efficiency bounds should be obtainable in the special case of conditional logit assumption.

Appendix A

In this section, we provide a set of high level assumptions (A1 - A6) and their consequences (C1 - C4) of the nonparametric estimators described in Section 3. We outline the stochastic expansions required to obtain the asymptotic properties of $\widehat{m}_\theta$ and $\widehat{g}_\theta$. The high level assumptions are then proved under the primitives of M1 - M3 and B1 - B8. Consequences are simple and their proofs are omitted. In what follows, we refer frequently to Bosq (1998), Linton and Mammen (2005), Masry (1996) and Robinson (1983), so for brevity, we denote their references by [B], [LM], [M] and [R] respectively.

A.1 Outline of Asymptotic Approach

For notational simplicity we work on a Banach space, $(C(\mathcal{X}), \|\cdot\|)$, where $\mathcal{X} = \mathcal{X}^C \times X^D$,the continuous part of $\mathcal{X}$ is a compact set $[\underline{x}+\varepsilon, \overline{x}-\varepsilon]$ for some arbitrarily small $\varepsilon > 0$. We denote $B1'$, the analogous condition to B1 when we replace X by $\mathcal{X}$. The approach taken here is similar to [LM], who worked on the L^2 Hilbert Space. The main difference between our problem and theirs is, after getting consistent estimates of (10), we require another level of smoothing (11) before plugging it into the criterion function. The first part here follows [LM].

ASSUMPTION A1. Suppose that for some sequence $\delta_T = o(1)$:

$$\sup_{x\in\mathcal{X}} \left| \left(\widehat{\mathcal{L}} - \mathcal{L}\right) m(x) \right| = o_p(\delta_T),$$

i.e., we have,

$$\left\| \left(\widehat{\mathcal{L}} - \mathcal{L}\right) m \right\| = o_p(\delta_T),$$

for any $m \in C(\mathcal{X})$.

CONSEQUENCE C1. Under A1:

$$\left\| \left(\left(I - \widehat{\mathcal{L}}\right)^{-1} - (I-\mathcal{L})^{-1} \right) m \right\| = o_p(\delta_T).$$

The rate of uniform approximation of the linear operator gets transferred to the inverse of $(I - \mathcal{L})$. This is summarized by C1 and is proven in [LM].

We supposed that $\widehat{r}_\theta(x) - r_\theta(x)$ can be decomposed into the following terms with some properties.

ASSUMPTION A2. For each $x \in$X:

$$\widehat{r}_\theta(x) - r_\theta(x) = \widehat{r}_\theta^B(x) + \widehat{r}_\theta^C(x) + \widehat{r}_\theta^D(x), \tag{25}$$

where $\widehat{r}_\theta^C, \widehat{r}_\theta^D$ and $\widehat{r}_\theta^E$ satisfy:

$$\begin{aligned}
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^B(x)\right| &= O_p\left(T^{-2/5}\right) \text{ with } \widehat{r}_\theta^B \text{ deterministic,} &(26)\\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^C(x)\right| &= o_p\left(T^{-2/5+\xi}\right) \text{ for any } \xi>0, &(27)\\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\mathcal{L}(I-\mathcal{L})^{-1}\widehat{r}_\theta^C(x)\right| &= o_p\left(T^{-2/5}\right), &(28)\\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^D(x)\right| &= o_p\left(T^{-2/5}\right). &(29)
\end{aligned}$$

This is the standard bias+variance+remainder of local constant kernel estimates of the regression function under some smoothness assumptions. The intuition behind (28), as provided in [LM], is that the operator applies averaging to a local smoother and transforms it into a global average thereby reducing its variance. These terms are used to obtain the components of $\widehat{m}_\theta(x)$, for $j=B,C,D$, the terms $\widehat{m}_\theta^j(x)$ are solutions to the integral equations,

$$\widehat{m}_\theta^j = \widehat{r}_\theta^j + \widehat{\mathcal{L}}\widehat{m}_\theta^j \tag{30}$$

and $\widehat{m}_\theta^A$, from writing the solution $m_\theta + \widehat{m}_\theta^A$ to the integral equation

$$\left(m_\theta + \widehat{m}_\theta^A\right) = r_\theta + \widehat{\mathcal{L}}\left(m_\theta + \widehat{m}_\theta^A\right). \tag{31}$$

The existence and uniqueness of the solutions to (30) and (31) are assured, at least w.p.a. 1, under the contraction property of the integral operator, so it follows from the linearity of $\left(I-\widehat{\mathcal{L}}\right)^{-1}$ that

$$\widehat{m}_\theta = m_\theta + \widehat{m}_\theta^A + \widehat{m}_\theta^B + \widehat{m}_\theta^C + \widehat{m}_\theta^D.$$

These components can be approximated by simpler terms. Define also m_θ^B, as the solution to

$$m_\theta^B = \widehat{r}_\theta^B + \mathcal{L}m_\theta^B. \tag{32}$$

Consequence C2. Under A1 - A2:

$$\begin{aligned}
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^B(x) - m_\theta^B(x)\right| &= o_p\left(T^{-2/5}\right), &(33)\\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^C(x) - r_\theta^C(x)\right| &= o_p\left(T^{-2/5}\right), &(34)\\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^D(x)\right| &= o_p\left(T^{-2/5}\right). &(35)
\end{aligned}$$

(33) and (35) follow immediately from (26), (29) and C1. (34) follows from (28), A1 and C1, as we can easily show that,

$$\left\|\widehat{\mathcal{L}}\left(I-\widehat{\mathcal{L}}\right)^{-1} - \mathcal{L}(I-\mathcal{L})^{-1}\right\| = o_p(\delta_T).$$

We next, also, approximate $\widehat{m}_\theta^A$ by simpler terms, subtracting (10) from (31) yields

$$\widehat{m}_\theta^A = \left(\widehat{\mathcal{L}} - \mathcal{L}\right) m_\theta + \widehat{\mathcal{L}} \widehat{m}_\theta^A. \tag{36}$$

Assumption A3. For any $x \in \mathcal{X}$:

$$\left(\widehat{\mathcal{L}} - \mathcal{L}\right) m_\theta (x) = \widehat{r}_\theta^E (x) + \widehat{r}_\theta^F (x) + \widehat{r}_\theta^G (x),$$

where $\widehat{r}_\theta^E, \widehat{r}_\theta^F$ and $\widehat{r}_\theta^G$ satisfy:

$$\begin{aligned}
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^E (x)\right| &= O_p\left(T^{-2/5}\right) \text{ with } \widehat{r}_\theta^E \text{ deterministic,} \\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^F (x)\right| &= o_p\left(T^{-2/5+\xi}\right) \text{ for any } \xi > 0, \\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\mathcal{L}\left(I - \mathcal{L}\right)^{-1} \widehat{r}_\theta^F (x)\right| &= o_p\left(T^{-2/5}\right), \\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{r}_\theta^G (x)\right| &= o_p\left(T^{-2/5}\right).
\end{aligned}$$

These terms are obtained by decomposing the conditional density estimates (cf. A2), then proceed as done previously, we define $\widehat{m}_\theta^j (x)$ for $j = E, F, G$ as the unique solutions to the estimated integral equation of (30), solving (36) we have,

$$\begin{aligned}
\widehat{m}_\theta^A &= \left(I - \widehat{\mathcal{L}}\right)^{-1} \left(\widehat{\mathcal{L}} - \mathcal{L}\right) m_\theta \\
&= \widehat{m}_\theta^E + \widehat{m}_\theta^F + \widehat{m}_\theta^G.
\end{aligned}$$

Such terms are asymptotically equivalent to more convenient terms (cf. C2), define also m_θ^E as the solution to the analogous integral equation of (32).

Consequence C3. Under A1 - A3:

$$\begin{aligned}
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^E (x) - m_\theta^E (x)\right| &= o_p\left(T^{-2/5}\right), \\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^F (x) - \widehat{r}_\theta^F (x)\right| &= o_p\left(T^{-2/5}\right), \\
\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta^G (x)\right| &= o_p\left(T^{-2/5}\right).
\end{aligned}$$

C3 can be shown using the same reasonings used to obtain C2.

Combining these assumptions leads to Proposition 1 of [LM].

Proposition 1. Suppose that [A1 - A3] holds for some estimators $\widehat{r}_\theta$ and $\widehat{\mathcal{L}}$. Define $\widehat{m}_\theta$ as any solution of $\widehat{m}_\theta = \widehat{r}_\theta + \widehat{\mathcal{L}} \widehat{m}_\theta$. Then the following expansion holds for $\widehat{m}_\theta$

$$\sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left|\widehat{m}_\theta (x) - m_\theta (x) - m_\theta^B (x) - m_\theta^E (x) - \widehat{r}_\theta^C (x) - \widehat{r}_\theta^F (x)\right| = o_p\left(T^{-2/5}\right),$$

where all of the terms above have been defined previously.

The uniform expansion for the nonparametric estimators required in [LM] ends here. However, to obtain the uniform expansion of $\widehat{g}_\theta$ defined in (19), we need another level of smoothing. Note that the integral operator, $\mathcal{H}$, has a different range,

$$\mathcal{H} : C(\mathcal{X}) \to C(A \times X),$$

where $C(A \times X)$ denotes a space of functions, say $g(j,x)$, which are continuous on X for each $j \in A$. So the relevant Banach Space is equipped with the sup-norm over $A \times X$, which we also denote by $\|\cdot\|$ though this should not lead to any confusion. For notational simplicity, we first define,

$$\begin{aligned}
\overline{m}_\theta^B(x) &= m_\theta^B(x) + m_\theta^E(x), \\
\overline{m}_\theta^C(x) &= \widehat{r}_\theta^C(x) + \widehat{r}_\theta^F(x), \\
\overline{m}_\theta^D(x) &= \widehat{m}_\theta(x) - m_\theta(x) - \overline{m}_\theta^B(x) - \overline{m}_\theta^C(x).
\end{aligned}$$

We next define various components of the transformations (19), analogously to (30) and (32), for $j = B, C, D$ the terms $\widehat{g}_\theta^j$ are elements of the integral transform,

$$\begin{aligned}
\widehat{g}_\theta^j &= \widehat{\mathcal{H}}\overline{m}_\theta^j, \\
g_\theta^j &= \mathcal{H}\overline{m}_\theta^j,
\end{aligned}$$

and $\widehat{g}_\theta^A$ is defined by

$$\widehat{\mathcal{H}} m_\theta = g_\theta + \widehat{g}_\theta^A.$$

It follows from linearity of $\widehat{\mathcal{H}}$ that

$$\widehat{g}_\theta = g_\theta + \widehat{g}_\theta^A + \widehat{g}_\theta^B + \widehat{g}_\theta^C + \widehat{g}_\theta^D.$$

Assumption A4. Suppose that for some sequence δ_T as in A1:

$$\sup_{(x,a,\theta)\in A\times X\times\Theta} \left| \left(\widehat{\mathcal{H}} - \mathcal{H}\right) m_\theta(x,a) \right| = o_p(\delta_T),$$

i.e., $\left\| \left(\widehat{\mathcal{H}} - \mathcal{H}\right) m \right\| = o_p(\delta_T)$ for any $m \in C(\mathcal{X})$.

A4 assumes the desirable properties of the conditional density estimators (cf. A1 and A3).

Consequence C4. Under A1 - A4:

$$\begin{aligned}
\sup_{(x,a,\theta)\in A\times X\times\Theta} \left| \widehat{g}_\theta^B(a,x) - g_\theta^B(a,x) \right| &= o_p\left(T^{-2/5}\right), \\
\sup_{(x,a,\theta)\in A\times X\times\Theta} \left| \widehat{g}_\theta^C(a,x) - g_\theta^C(a,x) \right| &= o_p\left(T^{-2/5}\right), \\
\sup_{(x,a,\theta)\in A\times X\times\Theta} \left| \widehat{g}_\theta^D(a,x) \right| &= o_p\left(T^{-2/5}\right).
\end{aligned}$$

This follows immediately from A5 and the properties of the elements defined in $\overline{m}_\theta^B(x)$.

ASSUMPTION A5. Suppose that:

$$\sup_{(x,a,\theta)\in A\times X\times\Theta}\left|g_\theta^C(a,x)\right| = o_p\left(T^{-2/5}\right).$$

A5 follows since the operator $\mathcal{H}$ is a global smooth, hence it reduces the variance of g_θ^C.

As with $\widehat{m}_\theta^A$ we can approximate $\widehat{g}_\theta^A$ by simpler terms.

ASSUMPTION A6. For any $m \in C(\mathcal{X})$ and for each $(a,x) \in A \times X$:

$$\begin{aligned}\widehat{g}_\theta^A(a,x) &= \left(\widehat{\mathcal{H}} - \mathcal{H}\right) m_\theta(x,j) \\ &= \widehat{g}_\theta^E(j,x) + \widehat{g}_\theta^F(j,x) + \widehat{g}_\theta^G(j,x),\end{aligned}$$

where $\widehat{g}_\theta^E, \widehat{g}_\theta^F$ and $\widehat{g}_\theta^G$ satisfy:

$$\begin{aligned}\sup_{(x,a,\theta)\in A\times X\times\Theta}\left|\widehat{g}_\theta^E(x,a)\right| &= O_p\left(T^{-2/5}\right) \text{ with } \widehat{g}_\theta^E \text{ deterministic,} \\ \sup_{(x,a,\theta)\in A\times X\times\Theta}\left|\widehat{g}_\theta^F(x,a)\right| &= o_p\left(T^{-2/5+\xi}\right) \text{ for any } \xi > 0, \\ \sup_{(x,a,\theta)\in A\times X\times\Theta}\left|\widehat{g}_\theta^G(x,a)\right| &= o_p\left(T^{-2/5}\right).\end{aligned}$$

A6 follows from standard decomposition of the kernel conditional density estimator (cf. A3).

PROPOSITION 2. Suppose that A1 - A6 holds for some estimators $\widehat{r}_\theta$, $\widehat{\mathcal{L}}$ and $\widehat{\mathcal{H}}$. Define $\widehat{m}_\theta$ as any solution of $\widehat{m}_\theta = \widehat{r}_\theta + \widehat{\mathcal{L}}\widehat{m}_\theta$ and $\widehat{g}_\theta = \widehat{\mathcal{H}}\widehat{m}_\theta$. Then the following expansion holds for $\widehat{g}_\theta$

$$\sup_{(x,a,\theta)\in A\times X\times\Theta}\left|\widehat{g}_\theta(x,a) - g_\theta(x,a) - g_\theta^B(x,a) - \widehat{g}_\theta^E(x,a) - \widehat{g}_\theta^F(x,a)\right| = o_p\left(T^{-2/5}\right),$$

where all of the terms above have been defined previously, in particular g_θ^B and $\widehat{g}_\theta^E$ are non-stochastic and the leading variance terms is $\widehat{g}_\theta^F$. This can be rewritten in a similar notation to (**??**).

$$\begin{aligned}\overline{g}_\theta^B(x,a) &= g_\theta^B(x,a) + \widehat{g}_\theta^E(x,a), \\ \overline{g}_\theta^C(x,a) &= \widehat{g}_\theta^F(x,a), \\ \overline{g}_\theta^D(x,a) &= \widehat{g}_\theta(x,a) - g_\theta(x,a) - \overline{g}_\theta^B(x,a) - \overline{g}_\theta^C(x,a).\end{aligned}$$

A.2 Proofs of Theorems 1 and 2 and High Level Conditions A1 - A6

We assume $B1'$ and B2 - B6 throughout this subsection. Set $\delta_T = T^{\xi-3/10}$, this rate is arbitrarily close to the rate of convergence of 1-dimensional nonparametric density estimates when h_T decays at the rate specified by B6. For the ease of notation, we assume that X^D is empty. The presence of discrete states do not affect any of the results below, we can simply replace any formula involving the density (and analogously for the conditional density) $f(dx_t)$ by $f\left(dx_t, x_t^d\right)$. We shall denote generic constants by C_0 that may take different values in different places. The uniform rate of convergence proof of various components utilize some exponential inequalities found in [B] as done in [LM], the details are deferred to Appendix B.

PROOF OF THEOREM 1. We proceed by providing the pointwise distribution theory for $\widehat{P}(a|x)$, for any $a \in A$ and $x \in \text{int}(X)$, and the functionals thereof. These are used to proof Theorem 1 and 2 and verify the high level conditions. $\widehat{P}(a|x)$ is the usual local constant regression estimator (or equivalently, the conditional probability estimator).

$$\widehat{P}(a|x) - P(a|x) = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{1}[a_t = a] - P(a|x)\right)K_h(x_t - x)/\widehat{f}_X(x),$$

focusing on the numerator

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{1}[a_t = a] - P(a|x)\right)K_h(x_t - x) &= \frac{1}{T}\sum_{t=1}^{T}\left(P(a|x_t) - P(a|x)\right)K_h(x_t - x) \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}e_{a,t}K_h(x_t - x) \\ &= A_{1,a,T}(x) + A_{2,a,T}(x),\end{aligned}$$

where $e_{a,t} = \mathbf{1}[a_t = a] - P(a|x_t)$. The term $A_{1,a,T}(x)$ is dominated by the bias, by the usual change of variables and Taylor's expansion,

$$\begin{aligned}E\left[A_{1,a,T}(x)\right] &= E\left[\left(P(a|x_t) - P(a|x)\right)K_h(x_t - x)\right] \\ &= \frac{1}{2}\mu_2 h_T^2\left(2\frac{\partial P(a|x)}{\partial x}\frac{\partial f_X(x)}{\partial x} + \frac{\partial^2 P(a|x)}{\partial x^2}f_X(x)\right) + o\left(h_T^2\right).\end{aligned}$$

Recall that $E\left[e_{a,t}|x_t\right] = 0$ for all a and t. We next compute the variance of $A_{2,a,T}(x)$, this is dominated by the variances as covariance terms are of smaller order, e.g. see [M].

$$\begin{aligned}\text{var}\left(A_{2,a,T}(x)\right) &= \text{var}\left(\frac{1}{T}\sum_{t=1}^{T}e_{a,t}K_h(x_t - x)\right) \\ &= \frac{1}{T}\text{var}\left(e_{a,t}K_h(x_t - x)\right) + o\left(\frac{1}{Th_T}\right) \\ &= \frac{1}{T}E\left[\sigma_a^2(x_t)K_h(x_t - x)\right] + o\left(\frac{1}{Th_T}\right) \\ &= \frac{\kappa_2}{Th_T}\sigma_a^2(x)f_X(x) + o\left(\frac{1}{Th_T}\right),\end{aligned}$$

note that

$$\begin{aligned}\sigma_a^2(x) &= E\left[e_{a,t}^2 | x_t = x\right] \\ &= \mathrm{var}\left(\mathbf{1}\left[a_t = a\right] | x_t = x\right) \\ &= P(a|x)\left(1 - P(a|x)\right).\end{aligned}$$

For the CLT, Lemma 7.1 of [R] can be used repeated throughout this section, using Bernstein blocking technique we obtain,

$$\sqrt{Th_T}\left(\widehat{P}(a|x) - P(a|x) - \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\right) \Longrightarrow \mathcal{N}\left(0, \omega_{P_a}(x)\right),$$

where

$$\begin{aligned}\eta_{P_a}(x) &= 2\frac{\frac{\partial P(a|x)}{\partial x}\frac{\partial f_X(x)}{\partial x}}{f_X(x)} + \frac{\partial P^2(a|x)}{\partial x^2}, \\ \omega_{P_a}(x) &= \kappa_2 \frac{\sigma_a^2(x)}{f_X(x)}.\end{aligned} \tag{37}$$

For any $\theta \in \Theta$, recall from (14) and (15)

$$\widehat{r}_\theta(x) = \sum_{a \in A} \zeta_{x,a,\theta}\left(\widehat{P}(a|x)\right),$$

where,

$$\zeta_{x,a,\theta}(t) = t\left(u_\theta(a,x) + \log t\right) + \gamma,$$

by mean value theorem (MVT),

$$\begin{aligned}&\zeta_{x,a,\theta}\left(\widehat{P}(a|x)\right) - \zeta_{x,a,\theta}\left(P(a|x) + \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\right) \\ =\ &\zeta'_{x,a,\theta}\left(P(a|x)\right)\left(\widehat{P}(a|x) - P(a|x) - \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\right) + o_p(1),\end{aligned}$$

and

$$\zeta'_{x,a,\theta}(t) = u_\theta(a,x) + \log t + 1. \tag{38}$$

By using MVT again, we can approximate $\zeta_{x,a,\theta}\left(P(a|x) + \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\right)$ more conveniently as follows,

$$\zeta_{x,a,\theta}\left(P(a|x) + \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\right) = \zeta_{x,a,\theta}\left(P(a|x)\right) + \frac{1}{2}\mu_2 h_T^2 \eta_{P_a}(x)\,\zeta'_{x,a,\theta}\left(P(a|x)\right) + o_p\left(h_T^2\right).$$

To obtain the asymptotic distribution for $\widehat{r}_\theta(x)$, we now provide the joint distribution of $\left\{\widehat{P}(a|x)\right\}$. It follows immediately, following [R], from Cramér-Wold device that

$$\sqrt{Th_T}\begin{pmatrix}\widehat{P}(1|x)-P(1|x)-\frac{1}{2}\mu_2 h_T^2\eta_{P_1}(x)\\ \vdots \\ \widehat{P}(K|x)-P(K|x)-\frac{1}{2}\mu_2 h_T^2\eta_{P_K}(x)\end{pmatrix}$$

$$\Longrightarrow \mathcal{N}\left(0,\frac{\kappa_2}{f_X(x)}\begin{pmatrix}\sigma_1^2(x) & \sigma_{2,1}^2(x) & \cdots & \sigma_{K,1}^2(x)\\ \sigma_{1,2}^2(x) & \ddots & & \vdots\\ \vdots & & \ddots & \sigma_{K,K-1}^2(x)\\ \sigma_{1,K}^2(x) & \cdots & \sigma_{K-1,K}^2(x) & \sigma_K^2(x)\end{pmatrix}\right),$$

where $\sigma_j^2(x)=P(j|x)(1-P(j|x))$ and $\sigma_{j,k}^2(x)=\sigma_{k,j}^2(x)=-P(j|x)P(k|x)$ for $j,k\in A$. There are a couple of things to notice here, first there exist negative correlation between $\left\{\widehat{P}(j|x)\right\}$ across A, and the covariance matrix in the above display is rank deficient due to the constraint that $\sum_{j\in A}\widehat{P}(j|x)=1$ for any $x\in\operatorname{int}(X)$. Using the information from the display above, we have

$$\sqrt{Th_T}\left(\widehat{r}_\theta(x)-r_\theta(x)-\frac{1}{2}\mu_2 h_T^2\eta_{r,\theta}\right)\Longrightarrow\mathcal{N}\left(0,\omega_{\zeta_x,\theta}(x)\right),$$

where

$$\eta_{r,\theta}(x) = \sum_{j\in A}\eta_{P_j}(x)\,\zeta'_{x,j,\theta}(P(j|x)), \tag{39}$$

$$\omega_{r,\theta}(x) = \frac{\kappa_2}{f_X(x)}\begin{pmatrix}\sum_{j\in A}\left(\zeta'_{x,j,\theta}(P(j|x))\right)^2\sigma_j^2(x)\\ -2\sum_{j\neq k}\zeta'_{x,j,\theta}(P(j|x))\,\zeta'_{x,k,\theta}(P(k|x))\,P(j|x)\,P(k|x)\end{pmatrix}, \tag{40}$$

where $\left\{\eta_{P_j}\right\}_{j\in A}$ and $\left\{\zeta'_{x,j,\theta}\right\}_{j\in A}$ are defined in (39) and (38) respectively. Note we can relate components of the expansion of $\widehat{r}_\theta(x)$, in (25), to the terms above as follows,

$$r_\theta(x) = \sum_{j\in A}\zeta_{x,j,\theta}(P(j|x)), \tag{41}$$

$$\widehat{r}_\theta^B(x) = \frac{1}{2}\mu_2 h_T^2\eta_{r,\theta}(x), \tag{42}$$

$$\widehat{r}_\theta^C(x) = \sum_{j\in A}\frac{\zeta'_{x,j,\theta}(P(j|x))}{f_X(x)}\times\left(\frac{1}{T}\sum_{t=1}^{T}e_{j,t}K_h(x_t-x)\right). \tag{43}$$

We next provide the statistical properties for $\widehat{m}^A_\theta(x)$. First, $\left(\widehat{\mathcal{L}}-\mathcal{L}\right)m_\theta(x)$:

$$
\begin{aligned}
\left(\widehat{\mathcal{L}}-\mathcal{L}\right)m_\theta(x) &= \beta\int m_\theta(x')\left(\widehat{f}_{X'|X}(dx'|x)-f_{X'|X}(dx'|x)\right)\\
&= \frac{\beta}{f_X(x)}\int m_\theta(x')\left(\widehat{f}_{X,X}(dx',x)-f_{X,X}(dx',x)\right)\\
&\quad -\frac{\beta}{f_X(x)}\left(\widehat{f}_X(x)-f_X(x)\right)\int m_\theta(x')f_{X'|X}(dx'|x)+o_p\left(T^{-2/5}\right)\\
&= B_{1,\theta,T}(x)+B_{2,\theta,T}(x)+o_p\left(T^{-2/5}\right).
\end{aligned}
$$

To analyze $B_{1,\theta,T}(x)$, proceed with the usual decomposition of $\widehat{f}_{X',X}(x',x)-f_{X',X}(x',x)$ then integrate it over, note that the integral reduces the variance to that of a 1 dimensional nonparametric estimator, we have

$$B_{1,\theta,T}(x)=B^B_{1,\theta,T}(x)+B^C_{1,\theta,T}(x)+o_p\left(T^{-2/5}\right),$$

where

$$
\begin{aligned}
B^B_{1,\theta,T}(x) &= \frac{1}{2}\mu_2h_T^2\beta\int\left(\frac{m_\theta(x')}{f_X(x)}\left(\frac{\partial^2 f_{X',X}(x',x)}{\partial x'^2}+\frac{\partial^2 f_{X',X}(x',x)}{\partial x^2}\right)\right)dx', &(44)\\
B^C_{1,\theta,T}(x) &= \frac{\beta}{f_X(x)}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\int m_\theta(x')\left(\begin{array}{c}K_h(x_{t+1}-x')K_h(x_t-x)\\-E\left[K_h(x_{t+1}-x')K_h(x_t-x)\right]\end{array}\right)\right)dx', &(45)
\end{aligned}
$$

and it can be shown that

$$\sqrt{Th_T}B^C_{1,\theta,T}(x)\Longrightarrow\mathcal{N}\left(0,\frac{\beta^2}{f_X(x)}\kappa_2\int\left(m_\theta(x')\right)^2f_{X'|X}(dx'|x)\right).$$

For $B_{2,\theta,T}(x)$, this is just the kernel density estimator of $f_X(x)$ multiplied by a non-stochastic term,

$$B_{2,\theta,T}(x)=B^B_{2,\theta,T}(x)+B^C_{2,\theta,T}(x)+o_p\left(T^{-2/5}\right),$$

where

$$
\begin{aligned}
B^B_{2,\theta,T}(x) &= -\frac{1}{2}\mu_2h_T^2\frac{\partial^2 f_X(x)}{\partial x^2}\left(\frac{\beta}{f_X(x)}\int m_\theta(x')f_{X'|X}(dx'|x)\right), &(46)\\
B^C_{2,\theta,T}(x) &= -\left(\frac{\beta}{f_X(x)}\int m_\theta(x')f_{X'|X}(dx'|x)\right)\frac{1}{T}\sum_{t=1}^{T}\left(K_h(x_t-x)-E\left[K_h(x_t-x)\right]\right), &(47)
\end{aligned}
$$

and it can be shown that

$$\sqrt{Th_T}B^C_{2,\theta,T}(x)\Longrightarrow\mathcal{N}\left(0,\kappa_2f_X(x)\left(\frac{\beta}{f_X(x)}\int m_\theta(x')f_{X'|X}(dx'|x)\right)^2\right).$$

Combining these we have,

$$\widehat{m}_\theta(x)=m_\theta(x)+\overline{m}^B_\theta(x)+\overline{m}^C_\theta(x)+o_p\left(T^{-2/5}\right),$$

where

$$\begin{aligned}\overline{m}_{\theta}^{B}(x) &= (I-\mathcal{L})^{-1}\left(B_{1,\theta,T}^{B}+B_{2,\theta,T}^{B}+\widehat{r}_{\theta}^{B}\right)(x),\\ \overline{m}_{\theta}^{C}(x) &= B_{1,\theta,T}^{C}(x)+B_{2,\theta,T}^{C}(x)+\widehat{r}_{\theta}^{C}(x).\end{aligned}$$

Note also that

$$\begin{aligned}\sqrt{Th_T}\left(B_{1,\theta,T}^{C}(x)+B_{2,\theta,T}^{C}(x)\right) &\Longrightarrow \mathcal{N}\left(0,\frac{\kappa_2\beta^2}{f_X(x)}\mathrm{var}\left(m_\theta(x_{t+1})\,|x_t=x\right)\right),\\ &\quad\text{and}\\ Cov\left(\sqrt{Th_T}\left(B_{1,\theta,T}^{C}(x)+B_{2,\theta,T}^{C}(x)\right),\sqrt{Th_T}\widehat{r}_{\theta}^{C}(x)\right) &\rightarrow 0\ \text{ as }\ n\rightarrow\infty.\end{aligned}$$

This provides us with the pointwise theory for $\widehat{m}_\theta$ for any $x\in\mathrm{int}(X)$ and $\theta\in\Theta$.

$$\sqrt{Th_T}\left(\widehat{m}_\theta(x)-m_\theta(x)-\frac{1}{2}\mu_2h_T^2\eta_{m,\theta}(x)\right)\Longrightarrow\mathcal{N}\left(0,\omega_{m,\theta}(x)\right),$$

where

$$\begin{aligned}\eta_{m,\theta}(x) &= (I-\mathcal{L})^{-1}\left(\eta_{r,\theta}+\eta_{\mathcal{L},\theta}\right)(x),\\ \omega_{m,\theta}(x) &= \frac{\kappa_2}{f_X(x)}\left(\beta^2\mathrm{var}\left(m_\theta(x_{t+1})\,|x_t=x\right)+\omega_{r,\theta}(x)\right),\end{aligned}$$

where $\eta_{r,\theta}$ and $\omega_{r,\theta}$ are defined in (39) and (40), and

$$\eta_{\mathcal{L},\theta}(x)=\beta\begin{pmatrix}\frac{1}{f_X(x)}\int m_\theta(x')\left(\frac{\partial^2 f_{X',X}(x',x)}{\partial x'^2}+\frac{\partial^2 f_{X',X}(x',x)}{\partial x^2}\right)dx'\\ -\frac{\frac{\partial^2 f_X(x)}{\partial x^2}}{f_X(x)}\int m_\theta(x')\,f_{X'|X}(dx'|x)\end{pmatrix}, \tag{48}$$

$\eta_{r,\theta}$, $\omega_{r,\theta}$ are defined in (39) - (40). The proof of pairwise asymptotic independence across distinct x is obvious. ■

Proof of Theorem 2. From the decomposition from Theorem 1 we obtain the pointwise results for $\widehat{g}_\theta(a,x)$. Similarly to the decomposition of $\left(\widehat{\mathcal{L}}-\mathcal{L}\right)m_\theta(x)$, we have

$$\begin{aligned}\left(\widehat{\mathcal{H}}-\mathcal{H}\right)m_\theta(x,a) &= \int m_\theta(x')\left(\widehat{f}_{X'|X,A}(dx'|x,a)-f_{X'|X,A}(dx'|x,a)\right)\\ &= C_{1,\theta,T}(a,x)+C_{2,\theta,T}(a,x)+o_p\left(T^{-2/5}\right).\end{aligned}$$

The properties for $C_{1,\theta,T}$ and $C_{2,\theta,T}$ are closely related to that of $B_{1,\theta,T}$ and $B_{2,\theta,T}$.

$$C_{1,\theta,T}(a,x)=C_{1,\theta,T}^{B}(a,x)+C_{1,\theta,T}^{C}(a,x)+o_p\left(T^{-2/5}\right),$$

where

$$\begin{aligned}C_{1,\theta,T}^{B}(a,x) &= \frac{1}{2}\mu_2h_T^2\int\frac{m_\theta(x')}{f_{X,A}(x,a)}\left(\frac{\partial^2 f_{X',X,A}(x',x,a)}{\partial x'^2}+\frac{\partial^2 f_{X',X,A}(x',x,a)}{\partial x^2}\right)dx',\\ C_{1,\theta,T}^{C}(a,x) &= \frac{1}{f_{X,A}(x,a)}\int\left(\frac{1}{T-1}\sum_{t=1}^{T-1}m_\theta(x')\begin{pmatrix}K_h(x_{t+1}-x')\,K_h(x_t-x)\,\mathbf{1}[a_t=a]\\ -E\left[K_h(x_{t+1}-x')\,K_h(x_t-x)\,\mathbf{1}[a_t=a]\right]\end{pmatrix}\right)dx',\end{aligned}$$

and as in the case of $B^C_{1,\theta,T}$

$$\sqrt{Th_T}C^C_{1,\theta,T}(a,x) \Longrightarrow \mathcal{N}\left(0, \frac{\kappa_2}{f_{X,A}(x,a)}\int (m_\theta(x'))^2 f_{X'|X,A}(dx'|x,a)\right).$$

Similarly for $C_{2,\theta,T}$,

$$\begin{aligned}
C^B_{2,\theta,T}(a,x) &= -\frac{1}{2}\mu_2 h_T^2 \frac{\partial^2 f_{X,A}(x,a)}{\partial x^2}\left(\frac{1}{f_{X,A}(x,a)}\int m_\theta(x') f_{X'|X,A}(dx'|x,a)\right),\\
C^C_{2,\theta,T}(a,x) &= -\left(\frac{1}{f_{X,A}(x,a)}\int m_\theta(x') f_{X'|X,A}(dx'|x,a)\right)\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} K_h(x_t-x)\mathbf{1}[a_t=a]\\ -E[K_h(x_t-x)\mathbf{1}[a_t=a]]\end{pmatrix},
\end{aligned}$$

and

$$\sqrt{Th_T}C^C_{2,\theta,T}(a,x) \Longrightarrow \mathcal{N}\left(0, \frac{\kappa_2}{f_{X,A}(x,a)}\int (m_\theta(x'))^2 f_{X'|X,A}(dx'|x,a)\right).$$

Combining these we have,

$$\widehat{g}_\theta(x,a) = g_\theta(x,a) + \overline{g}^B_\theta(x,a) + \overline{g}^C_\theta(x,a) + o_p\left(T^{-2/5}\right),$$

where

$$\begin{aligned}
\overline{g}^B_\theta(x,a) &= C^B_{1,\theta,T}(a,x) + C^B_{2,\theta,T}(a,x) + \mathcal{H}\widehat{r}^B_\theta(a,x),\\
\overline{g}^C_\theta(x,a) &= C^C_{1,\theta,T}(a,x) + C^C_{2,\theta,T}(a,x).
\end{aligned}$$

This provides us with the pointwise distribution theory for $\widehat{g}$ for any $x \in \mathrm{int}(X)$, $j \in A$ and $\theta \in \Theta$.

$$\sqrt{Th_T}\left(\widehat{g}(x,a) - g(x,a) - \frac{1}{2}\mu_2 h_T^2 \eta_{g,\theta}(x,a)\right) \Longrightarrow \mathcal{N}(0, \omega_{g,\theta}(x,a)),$$

where,

$$\begin{aligned}
\eta_{g,\theta}(a,x) &= \mathcal{H}(I-\mathcal{L})^{-1}\left(\eta_{r,\theta}+\eta_{\mathcal{L},\theta}\right)(x,a) + \eta_{\mathcal{H},\theta}(x,a),\\
\omega_{g,\theta}(a,x) &= \frac{\kappa_2}{f_{X,A}(a,x)}\mathrm{var}\left(m_\theta(x_{t+1})|x_t=x, a_t=j\right),
\end{aligned}$$

$\eta_{r,\theta}$ and $\eta_{\mathcal{L},\theta}$ are as defined in the proof of Theorem 1, and

$$\begin{aligned}
\eta_{\mathcal{H},\theta}(a,x) &= \frac{1}{f_{X,A}(x,a)}\int m_\theta(x')\left(\frac{\partial^2 f_{X',X,A}(x',x,a)}{\partial^2 x'} + \frac{\partial^2 f_{X',X,A}(x',x,a)}{\partial^2 x}\right)dx' \qquad (49)\\
&\quad -\frac{\frac{\partial^2 f_{X,A}(x,a)}{\partial^2 x}}{f_{X,A}(x,a)}\int m_\theta(x') f_{X'|X,A}(dx'|x,a).
\end{aligned}$$

Pairwise asymptotic independence, across distinct x, completes the proof. ■

Proof of A1. It suffices to show that

$$\begin{aligned}\sup_{(x',x)\in\mathcal{X}\times\mathcal{X}}\left|\widehat{f}_{X',X}\left(x',x\right)-f_{X',X}\left(x',x\right)\right| &= o_p\left(\delta_T\right),\\ \sup_{x\in\mathcal{X}}\left|\widehat{f}_X\left(x\right)-f_X\left(x\right)\right| &= o_p\left(\delta_T\right).\end{aligned}$$

These uniform rates are bounded by the rates for the bias squared and the rates of the centred process. The former is standard, and holds uniformly over $\mathcal{X}\times\mathcal{X}$ (and $\mathcal{X}$). See Appendix B, where proof of A1 falls under Case 1. ■

Proof of A2. The components for the decomposition have been provided by (41) - (43). By uniform boundedness of η_{P_a} and $\zeta_{x,a,\theta}$ over $A\times X\times\Theta$ and triangle inequalities, the order of the leading bias and remainder terms are as stated in (26) and (29) respectively. For the stochastic term, we can utilize the exponential inequality, see Case 2 of Appendix B. We next check (28). [LM] use eigen-expansion to construct the kernel of the new integral operator and showed that it had nice properties in their problem. We use the Neumann's series to construct our kernel, for any $\phi\in C\left(\mathcal{X}\right)$

$$\mathcal{L}\left(I-\mathcal{L}\right)^{-1}\phi=\sum_{j=1}^{\infty}\mathcal{L}^j\phi, \tag{50}$$

where $\mathcal{L}^j$ represents a linear operator of a j-step ahead predictor with discounting, this follows from Chapman-Kolmogorov equation for homogeneous Markov chains, for $\tau>1$

$$\begin{aligned}\mathcal{L}^{\tau}\phi\left(x\right) &= \beta^{\tau}\int\phi\left(x'\right)f_{(\tau)}\left(dx'|x\right) \qquad (51)\\ f_{(\tau)}\left(x_{t+\tau}|x_t\right) &= \int f_{X'|X}\left(x_{t+\tau}|x_{t+\tau-1}\right)\prod_{k=1}^{\tau-1}f_{X'|X}\left(dx_{t+\tau-k}|x_{t+\tau-k-1}\right),\end{aligned}$$

where $f_{(\tau)}\left(dx_{t+\tau}|x_t\right)$ denotes the conditional density of τ-steps ahead. First, we note that $\mathcal{L}\left(I-\mathcal{L}\right)^{-1}\phi\in C\left(\mathcal{X}\right)$, this is always true since for any $\phi\in C\left(\mathcal{X}\right)$ and $x\in\mathcal{X}$ since:

$$\begin{aligned}\left|\mathcal{L}\left(I-\mathcal{L}\right)^{-1}\phi\left(x\right)\right| &= \left|\sum_{\tau=1}^{\infty}\beta^{\tau}\int\phi\left(x'\right)f_{(\tau)}\left(dx'|x\right)\right|\\ &\leq \sum_{\tau=1}^{\infty}\beta^{\tau}\int f_{(\tau)}\left(dx'|x\right)\left\|\phi\right\|\\ &\leq \frac{\beta}{1-\beta}\left\|\phi\right\|\\ &< \infty.\end{aligned}$$

We denote the kernel of the integral transform (50) by the limit, φ, of the partial sum, φ_J,

$$\varphi_{\mathcal{T}}\left(x',x\right)=\sum_{\tau=1}^{\mathcal{T}}\beta^{\tau}f_{(\tau)}\left(x'|x\right), \tag{52}$$

where φ is continuous on $\mathcal{X} \times \mathcal{X}$. This is easy to see since $f_{(\tau)}$ is continuous and is uniformly bounded for all j by $\sup_{(x',x)\in\mathcal{X}\times\mathcal{X}} |f(x'|x)|$, by completeness, φ_J converges to a continuous function (with Lipschitz constant no larger than $\frac{\beta}{1-\beta}\sup_{(x',x)\in\mathcal{X}\times\mathcal{X}} |f(x'|x)|$). To proof (28), for details see Case 3 of Appendix B, we apply exponential inequality to bound

$$\Pr\left(\frac{1}{T}\left|\sum_{t=1}^{T} e_{\theta,t}\nu(x_t,x)\right| > \delta_T\right), \tag{53}$$

for some positive sequence, $\delta_T = o\left(T^{-2/5}\right)$, where $\nu(x_t,x)$ is defined as

$$\begin{aligned}\nu(x_t,x) &= \int \frac{K_h(x_t-x')}{f_X(x')}\varphi(dx',x) \\ &= \frac{\varphi(x_t,x)}{f_X(x_t)} + O\left(h_T^2\right),\end{aligned} \tag{54}$$

and the latter equality holds uniformly on $\mathcal{X}$. ■

PROOF OF A3. Following the decomposition of $\widehat{f}(x'|x)$ we obtain the leading bias and variance terms are sum of (44) and (46), and, (45) and (47) respectively. The results rates of convergence follow similarly to the proof of A2. ■

PROOF OF A4. This is essentially the same as proof of A1. ■

PROOF OF A5. Notice that $\overline{m}_\theta^C$ consists of $\widehat{r}_\theta^C$ and $\widehat{r}_\theta^F$. We need to show,

$$\begin{aligned}\sup_{(j,x)\in A\times X}\left|\mathcal{H}\widehat{r}_\theta^C(j,x)\right| &= o_p\left(T^{-2/5}\right)\\ \sup_{(j,x)\in A\times X}\left|\mathcal{H}\widehat{r}_\theta^F(j,x)\right| &= o_p\left(T^{-2/5}\right).\end{aligned}$$

The proof follows from exponential inequalities, see Appendix B. ■

PROOF OF A6. This is essentially the same as proof of A3. ■

A.3 Proofs of Theorems 3 - 5

We begin with two lemmas for the uniform expansion of some partial derivatives of $\widehat{m}_\theta$ and $\widehat{g}_\theta$.

LEMMA 1: Under conditions $B1'$, B2 - B6 hold. Then the following expansion holds for $k = 0, 1, 2$ and $j = 1, \ldots, L$,

$$\max_{1\le j\le L}\sup_{(x,\theta)\in\mathcal{X}\times\Theta}\left|\frac{\partial^k\widehat{m}_\theta(x)}{\partial\theta_j^k} - \frac{\partial^k m_\theta(x)}{\partial\theta_j^k} - \frac{\partial^k\overline{m}_\theta^B(x)}{\partial\theta_j^k} - \frac{\partial^k\overline{m}_\theta^C(x)}{\partial\theta_j^k}\right| = o_p\left(T^{-2/5}\right),$$

where $\frac{\partial^k m_\theta}{\partial\theta^k}$ is defined as the solution to

$$\frac{\partial^k m_\theta}{\partial\theta_j^k} = \frac{\partial^k r_\theta}{\partial\theta_j^k} + \mathcal{L}\frac{\partial^k m_\theta}{\partial\theta_j^k}, \tag{55}$$

and $\frac{\partial^k \widehat{m}_\theta}{\partial \theta_j^k}$ defined as the solution to the analogous empirical integral equation. Standard definition for partial derivative applies for $\frac{\partial^k \overline{m}_\theta^b(x)}{\partial \theta_j^k}$ with $b = B, C$. Notice, when $k = 0$, this coincides with the terms previously defined in Proposition 1. Further,

$$\begin{aligned}
\max_{1\leq j\leq L} \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left| \frac{\partial^k \overline{m}_\theta^B(x)}{\partial \theta_j^k} \right| &= O_p\left(T^{-2/5}\right) \text{ with } \frac{\partial^k \overline{m}_\theta^B(x)}{\partial \theta^k} \text{ deterministic,} \\
\max_{1\leq j\leq L} \sup_{(x,\theta)\in\mathcal{X}\times\Theta} \left| \frac{\partial^k \overline{m}_\theta^C(x)}{\partial \theta_j^k} \right| &= o_p\left(T^{\xi-2/5}\right) \text{ for any } \xi > 0.
\end{aligned}$$

Proof of Lemma 1. Comparing integral equations in (10) and (55), we notice that, these are just the integral equations with the same kernel but different intercepts. Since $\zeta_{x,j,\theta}$, $\zeta_{x,j,\theta}$ and m_θ are twice continuously differentiable in θ on Θ over $A \times X$, Dominated Convergence Theorem (DCT) can be utilized throughout, all arguments used to verify the definition of $\frac{\partial^k \overline{m}_\theta(x)}{\partial \theta_j^k}$ and their uniformity results analogous to A2 -A3 follow immediately. ■

Lemma 2: Under conditions $B1'$, B2 - B6 hold. Then the following expansion holds for $k = 0, 1, 2$ and $j = 1, \ldots, L$,

$$\max_{1\leq j\leq L} \sup_{(x,j,\theta)\in A\times X\times\Theta} \left| \frac{\partial^k \widehat{g}_\theta(x,a)}{\partial \theta_j^k} - \frac{\partial^k g_\theta(x,a)}{\partial \theta_j^k} - \frac{\partial^k \overline{g}_\theta^B(x,a)}{\partial \theta_j^k} - \frac{\partial^k \overline{g}_\theta^C(x,a)}{\partial \theta_j^k} \right| = o_p\left(T^{-2/5}\right),$$

where all of the terms above are defined analogously to those found in Lemma 1 and, for $k = 1, 2$

$$\begin{aligned}
\max_{1\leq j\leq L} \sup_{(x,j,\theta)\in A\times X\times\Theta} \left| \frac{\partial^k \overline{g}_\theta^B(x,a)}{\partial \theta_j^k} \right| &= O_p\left(T^{-2/5}\right) \text{ with } \frac{\partial^k \overline{g}_\theta^B(x,a)}{\partial \theta_j^k} \text{ deterministic,} \\
\max_{1\leq j\leq L} \sup_{(x,j,\theta)\in A\times X\times\Theta} \left| \frac{\partial^k \overline{g}_\theta^C(x,a)}{\partial \theta_j^k} \right| &= o_p\left(T^{\xi-2/5}\right) \text{ for any } \xi > 0.
\end{aligned}$$

Proof of Lemma 2: Same as the proof of Lemma 1. ■

Proof of Theorem 3: We first proceed to show the consistency result of the estimator.

Consistency.

Consider any estimator θ_T of θ_0 that asymptotically maximizes $\widehat{Q}_T(\theta)$:

$$Q_T(\theta_T) \geq \sup_{\theta\in\Theta} Q_T(\theta) - o_p(1).$$

Under B1 and B9, by standard arguments for example see McFadden and Newey (1994), consistency of such extremum estimators can be obtained if we have

$$\sup_{\theta\in\Theta} \left| \widehat{Q}_T(\theta) - Q(\theta) \right| = o_p(1). \tag{56}$$

By triangle inequality, (56) is implied by

$$\sup_{\theta\in\Theta}\left|Q_T(\theta)-Q(\theta)\right| = o_p(1) \quad (57)$$

$$\sup_{\theta\in\Theta}\left|\widehat{Q}_T(\theta)-Q_T(\theta)\right| = o_p(1). \quad (58)$$

For (57), since $q: A\times X\times\Theta\to\mathbb{R}$ is continuous on the compact set $X\times\Theta$, for any $a\in A$, hence by Weierstrass Theorem

$$\max_{a\in A}\sup_{x\in X,\theta\in\Theta}\left|q(j,x;\theta,g_\theta)\right|<\infty. \quad (59)$$

This ensures that $E\left|q(a_t,x_t;\theta,v_\theta)\right|<\infty$, and by the LLN for ergodic and stationary processes we have

$$Q_T(\theta)\xrightarrow{p}Q(\theta) \quad \text{for each } \theta\in\Theta.$$

The convergence above can be made uniform since Q_T is stochastic equicontinuous and Q is uniformly continuous by DCT, with a majorant in (59). To proof (58) we partition $\widehat{Q}_T(\theta)-Q_T(\theta)$ into two components

$$\widehat{Q}_T(\theta)-Q_T(\theta)=\frac{1}{T}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\left(q(a_t,x_t;\theta,\widehat{g}_\theta)-q(a_t,x_t;\theta,g_\theta)\right)+\frac{1}{T}\sum_{t=1}^{T}\left(1-\mathfrak{c}_{t,T}\right)q(a_t,x_t;\theta,\widehat{g}_\theta),$$

where the second term is $o_p(1)$. This follows since, we denote $1-\mathfrak{c}_{t,T}$ by $\mathfrak{d}_{t,T}$, ,

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}\mathfrak{d}_{t,T}q(a_t,x_t;\theta,\widehat{g}_\theta)\right| &\leq \max_{a\in A}\sup_{x\in X,\theta\in\Theta}\left|q(a,x;\theta,g_\theta)\right|\frac{1}{T}\sum_{t=1}^{T}\mathfrak{d}_{t,T}\\ &= o_p(1).\end{aligned}$$

The first inequality holds w.p.a. 1 and the equality is the result of $\mathfrak{d}_{t,T}=o_p(\vartheta_T)$ for any rate $\vartheta_T\to\infty$. To proof (58), now it suffices to show,

$$\max_{a\in A}\sup_{x\in X,\theta\in\Theta}\left|q(a,x;\theta,\widehat{g}_\theta)-q(a,x,j;\theta,g_\theta)\right|=o_p(1).$$

Recall that

$$\begin{aligned}q(j,x;\theta,\widehat{g}_\theta)-q(j,x;\theta,g_0) &= \widehat{v}_\theta(x,a)-v_\theta(x,a)+\log\left(\frac{\sum_{\widetilde{a}\in A}\exp\left(v_\theta(\widetilde{a},x)\right)}{\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_\theta(\widetilde{a},x)\right)}\right),\\ v_\theta(x,a) &= u_\theta(x,a)+g_\theta(x,a),\\ \widehat{v}_\theta(x,a) &= u_\theta(x,a)+\widehat{g}_\theta(x,a).\end{aligned}$$

All the listed functions are in $C(\mathcal{X})$. We have shown earlier that for some $\delta_T=o(1)$

$$\max_{a\in A}\sup_{x\in X,\theta\in\Theta}\left|\widehat{g}_\theta(x,a)-g_\theta(x,a)\right|=o_p(\delta_T),$$

so we have uniform convergence for $\widehat{v}$ to v at the same rate. We know for any continuously differentiable function ϕ (in this case, $\exp(\cdot)$ and $\log(\cdot)$), by MVT,

$$\max_{a\in A}\sup_{x\in X,\theta\in\Theta}\left|\phi\left(\widehat{v}_{\theta}\left(a,x\right)\right)-\phi\left(v_{\theta}\left(a,x\right)\right)\right|=o_{p}\left(\delta_{T}\right).$$

So we have

$$\sup_{x\in X,\theta\in\Theta}\left|\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)-\sum_{\widetilde{a}\in A}\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)\right|=o_{p}\left(1\right),$$

and since we have, at least w.p.a. 1, $\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)$ and $\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)$ are positive a.s.

$$\left|\frac{\sum_{\widetilde{a}\in A}\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)}{\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)}-1\right|=\left|\frac{1}{\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)}\right|\left|\sum_{\widetilde{a}\in A}\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)-\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)\right|,$$

and by Wierstrass Theorem, w.p.a. 1,

$$\min_{a\in A}\inf_{x\in X,\theta\in\Theta}\exp\left(\widehat{v}_{\theta}\left(a,x\right)\right)>0,$$

hence we have

$$\sup_{x\in X,\theta\in\Theta}\left|\frac{\sum_{\widetilde{a}\in A}\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)}{\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)}-1\right|=o_{p}\left(1\right).$$

The proof of (58) is completed once we apply another mean value expansion, as done previously, to obtain

$$\sup_{x\in X,\theta\in\Theta}\left|\log\left(\frac{\sum_{\widetilde{a}\in A}\exp\left(v_{\theta}\left(\widetilde{a},x\right)\right)}{\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta}\left(\widetilde{a},x\right)\right)}\right)\right|=o_{p}\left(1\right).$$

Asymptotic Normality

Consider the first order condition

$$\frac{\partial\widehat{Q}_{T}\left(\widehat{\theta}\right)}{\partial\theta}=o_{p}\left(1\right),$$

from MVT we have

$$o_{p}\left(\sqrt{T}\right)=\sqrt{T}\frac{\partial\widehat{Q}_{T}\left(\theta_{0}\right)}{\partial\theta}+\frac{\partial^{2}\widehat{Q}_{T}\left(\overline{\theta}\right)}{\partial\theta^{2}}\sqrt{T}\left(\widehat{\theta}-\theta_{0}\right).$$

We show that for any sequence $\epsilon_{T}\rightarrow0$ there exists some positive C such that

$$\inf_{\|\theta-\theta_{0}\|<\epsilon_{T}}\lambda_{\min}\left(-\frac{\partial^{2}\widehat{Q}_{T}\left(\theta\right)}{\partial\theta\partial\theta^{\top}}\right)>C+o_{p}\left(1\right)\tag{60}$$

$$\sqrt{T}\frac{\partial\widehat{Q}_{T}\left(\theta_{0}\right)}{\partial\theta}=O_{p}\left(1\right)\tag{61}$$

This implies

$$\sqrt{T}\left(\widehat{\theta}-\theta_{0}\right)=O_{p}\left(1\right).$$

To proof (60), we first show

$$\sup_{\|\theta-\theta_0\|<\epsilon_T}\left\|\frac{\partial^2\widehat{Q}_T(\theta)}{\partial\theta\partial\theta^\top}-E\left[\frac{\partial^2 q(a_t,x_t;\theta,g_\theta)}{\partial\theta\partial\theta^\top}\right]\right\|=o_p(1). \tag{62}$$

Since the second derivative of $q: A\times X\times\Theta\to\mathbb{R}$ is continuous on the compact set $X\times\Theta$ and for each $a\in A$, standard arguments for uniform convergence implies that

$$\sup_{\|\theta-\theta_0\|<\epsilon_T}\left\|\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta^\top}-E\left[\frac{\partial^2 q(a_t,x_t;\theta,g_\theta)}{\partial\theta\partial\theta^\top}\right]\right\|=o_p(1).$$

By triangle inequality, (62) will hold if we can show,

$$\sup_{\|\theta-\theta_0\|<\epsilon_T}\left\|\frac{\partial^2\widehat{Q}_T(\theta)}{\partial\theta\partial\theta^\top}-\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta^\top}\right\|=o_p(1).$$

This is similar to showing (58), as the above condition is implied by,

$$\max_{a\in A}\sup_{x\in X,\|\theta-\theta_0\|<\epsilon_T}\left\|\frac{\partial^2 q(a_t,x_t;\theta,\widehat{g}_\theta)}{\partial\theta\partial\theta^\top}-\frac{\partial^2 q(a_t,x_t;\theta,g_\theta)}{\partial\theta\partial\theta^\top}\right\|=o_p(1). \tag{63}$$

The expressions for the score of q is,

$$\frac{\partial q(a_t,x_t;\theta,g_\theta)}{\partial\theta}=\frac{\partial v_\theta(a_t,x_t)}{\partial\theta}-\frac{\sum_{\widetilde{a}\in A}\left(\frac{\partial v_\theta(\widetilde{a},x_t)}{\partial\theta}\right)\exp(v_\theta(\widetilde{a},x_t))}{\sum_{\widetilde{a}\in A}\exp(v_\theta(\widetilde{a},x_t))}, \tag{64}$$

and for the Hessian

$$\begin{aligned}\frac{\partial^2 q(a_t,x_t;\theta,g_\theta)}{\partial\theta\partial\theta^\top} &= \frac{\partial^2 v_\theta(a_t,x_t)}{\partial\theta\partial\theta^\top}-\frac{\sum_{\widetilde{a}\in A}\left(\frac{\partial^2 v_\theta(\widetilde{a},x_t)}{\partial\theta\partial\theta^\top}+\frac{\partial v_\theta(\widetilde{a},x_t)}{\partial\theta}\frac{\partial v_\theta(\widetilde{a},x_t)}{\partial\theta^\top}\right)\exp(v_\theta(\widetilde{a},x_t))}{\sum_{\widetilde{a}\in A}\exp(v_\theta(\widetilde{a},x_t))}\\ &\quad+\frac{\sum_{\widetilde{a}\in A}\frac{\partial v_\theta(\widetilde{a},x_t)}{\partial\theta}\frac{\partial v_\theta(\widetilde{a},x_t)}{\partial\theta^\top}\exp(v_\theta(\widetilde{a},x_t))^2}{\left(\sum_{\widetilde{a}\in A}\exp(v_\theta(\widetilde{a},x_t))\right)^2}.\end{aligned}$$

Proceed along the same line of arguments for proving (58), we show (63) holds by tedious but straightforward calculations. Essentially we need uniform convergence of the following partial derivatives,

$$\max_{a\in A,1\le j\le q}\sup_{x\in X,\theta\in\Theta}\left|\frac{\partial^k\widehat{v}_\theta(j,x)}{\partial\theta_j^k}-\frac{\partial^k v_\theta(j,x)}{\partial\theta_j^k}\right|=o_p(1)\quad\text{for}\quad k=0,1,2, \tag{65}$$

(63) follows from repeated mean value expansions as done in the proof of (58). The uniform convergence in (65) follows from Lemma 1 and 2, this implies (60).

For (61),

$$
\begin{aligned}
\sqrt{T}\frac{\partial \widehat{Q}_T(\theta_0)}{\partial \theta} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\frac{\partial q\left(a_t, x_t; \theta_0, \widehat{g}_{\theta_0}\right)}{\partial \theta} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta} \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\left(\frac{\partial q\left(a_t, x_t; \theta_0, \widehat{g}_{\theta_0}\right)}{\partial \theta} - \frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}\right) \\
&\quad \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{d}_{t,T}\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta} \\
&= D_{1,T} + D_{2,T} + D_{3,T},
\end{aligned}
$$

The term $D_{1,T}$ is asymptotically normal with mean zero and finite variance by the CLT for stationary and geometric mixing process,

$$
\sqrt{T}D_{1,T} \Longrightarrow \mathcal{N}\left(0, \Lambda_1\right),
$$

where

$$
\begin{aligned}
\Lambda_1 &= E\left[\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}^{\top}\right] \\
&\quad + \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(T-t)\left(\begin{array}{c} E\left[\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}\frac{\partial q\left(x_0, a_0; \theta_0, g_{\theta_0}\right)}{\partial \theta}^{\top}\right] \\ +E\left[\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}\frac{\partial q\left(x_0, a_0; \theta_0, g_{\theta_0}\right)}{\partial \theta}^{\top}\right]^{\top} \end{array}\right).
\end{aligned}
$$

Note that $E\left[\frac{\partial q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial \theta}\right] = 0$ by definition of θ_0. Next we show that $D_{2,T}$ also converges to a normal vector at the rate $1/\sqrt{T}$. Consider the j-th element of $D_{2,T}$, using the expression from the score function defined in (64),

$$
\begin{aligned}
\left(D_{2,T}\right)_j &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\left(\frac{\partial \widehat{v}_{\theta_0}\left(\widetilde{a}, x_t\right)}{\partial \theta_j} - \frac{\partial v_{\theta_0}\left(\widetilde{a}, x_t\right)}{\partial \theta_j}\right) \\
&\quad - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\left(\begin{array}{c} \left(\sum_{\widetilde{a}\in A}\frac{\partial \widehat{v}_{\theta_0}(\widetilde{a}, x_t)}{\partial \theta_j}\exp\left(\widehat{v}_{\theta_0}\left(\widetilde{a}, x_t\right)\right)\right) / \left(\sum_{\widetilde{a}\in A}\exp\left(\widehat{v}_{\theta_0}\left(\widetilde{a}, x_t\right)\right)\right) \\ -\left(\sum_{\widetilde{a}\in A}\frac{\partial v_{\theta_0}(\widetilde{a}, x_t)}{\partial \theta_j}\exp\left(v_{\theta_0}\left(\widetilde{a}, x_t\right)\right)\right) / \left(\sum_{\widetilde{a}\in A}\exp\left(v_{\theta_0}\left(\widetilde{a}, x_t\right)\right)\right) \end{array}\right),
\end{aligned}
$$

linearizing,

$$
\begin{aligned}
(D_{2,T})_j &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathfrak{c}_{t,T}\left(\frac{\partial\widehat{v}_{\theta_0}(a_t,x_t)}{\partial\theta_j}-\frac{\partial v_{\theta_0}(a_t,x_t)}{\partial\theta_j}\right)\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{\widetilde{a}\in A}\mathfrak{c}_{t,T}\psi_1(\widetilde{a},x_t)\left(\frac{\partial\widehat{v}_{\theta_0}(\widetilde{a},x_t)}{\partial\theta_j}-\frac{\partial v_{\theta_0}(\widetilde{a},x_t)}{\partial\theta_j}\right)\\
&\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{\widetilde{a}\in A}\mathfrak{c}_{t,T}\psi_{2,j}(\widetilde{a},x_t)\left(\widehat{v}_{\theta_0}(\widetilde{a},x_t)-v_{\theta_0}(\widetilde{a},x_t)\right)\\
&\quad +\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{\widetilde{a}\in A}\mathfrak{c}_{t,T}\psi_{2,j}(\widetilde{a},x_t)\left(\sum_{\widetilde{\widetilde{a}}\in A}P\left(\widetilde{\widetilde{a}}|x_t\right)\left(\widehat{v}_{\theta_0}\left(\widetilde{\widetilde{a}},x_t\right)-v_{\theta_0}\left(\widetilde{\widetilde{a}},x_t\right)\right)\right)+o_p(1)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(E_{1,t,T})_j+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(E_{2,t,T})_j+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(E_{3,t,T})_j+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(E_{4,t,T})_j+o_p(1),
\end{aligned}
$$

where

$$
\psi_1(\widetilde{a},x_t) = P(\widetilde{a}|x_t), \tag{66}
$$

$$
\psi_{2,j}(\widetilde{a},x_t) = P(\widetilde{a}|x_t)\frac{\partial v_{\theta_0}(\widetilde{a},x_t)}{\partial\theta_j}, \tag{67}
$$

and the remainder terms are of smaller order since our nonparametric estimates converge uniformly to the true at the rate faster than $T^{-1/4}$ on the trimming set, as proven in Theorem 1 and 2.

The asymptotic properties of these terms are tedious but simple to obtain. We utilize the projection results and law of large numbers for U-statistics, see Lee (1990). We also note that all of the relevant kernels for our statistics are uniformly bounded, along with the assumption [B1], this ensures the residuals from the projections can be ignored. Now we give some details for deriving the distribution of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(E_{1,t,T})_j$. First we linearize $\frac{\partial^k\widehat{g}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k g_{\theta_0}}{\partial\theta_j^k}$ for $k=0,1$,

$$
\begin{aligned}
\frac{\partial^k\widehat{g}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k g_{\theta_0}}{\partial\theta_j^k} &= \widehat{\mathcal{H}}\frac{\partial^k\widehat{m}_{\theta_0}}{\partial\theta_j^k}-\mathcal{H}\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}\\
&\sim \left(\widehat{\mathcal{H}}-\mathcal{H}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}+\mathcal{H}(I-\mathcal{L})^{-1}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}\\
&\quad +\mathcal{H}(I-\mathcal{L})^{-1}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right),
\end{aligned}
$$

this expansion is valid, uniformly on the trimming set, in spite of the scaling of order $\sqrt{T}$. Consider the normalized sum of $\left(\widehat{\mathcal{H}}-\mathcal{H}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}$, with further linearization, see the decomposition $\widehat{\mathcal{L}}-\mathcal{L}$ and

$\widehat{\mathcal{H}}-\mathcal{H}$ in the proof of [A1], we can obtain the following U-statistics, scaled by $\sqrt{T}$, representation,

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathfrak{c}_{t,T}\left[\left(\widehat{\mathcal{H}}-\mathcal{H}\right)\frac{\partial m_{\theta_0}}{\partial\theta_j}(x_t,a_t)\right] \\
= & \frac{1}{\sqrt{T}}\frac{1}{T-1}\sum_{t=1}^{T-1}\sum_{s\neq t}\mathfrak{c}_{t,T}\left(\begin{array}{c}\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\frac{K_h(x_s-x_t)\mathbf{1}[a_s=a_t]}{f_{X,A}(x_t,a_t)}-E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t,a_t\right]\\ -\frac{K_h(x_s-x_t)\mathbf{1}[a_s=a_t]-f_{X,A}(x_t,a_t)}{f_{X,A}(x_t,a_t)}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t,a_t\right]\end{array}\right)+o_p(1) \\
= & \sqrt{T}\left(\begin{array}{c}T-1\\2\end{array}\right)^{-1}\sum_{t=1}^{T-1}\sum_{s>t}\frac{1}{2}\left(\begin{array}{c}\mathfrak{c}_{t,T}\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\frac{K_h(x_s-x_t)\mathbf{1}[a_s=a_t]}{f_{X,A}(x_t,a_t)}-\mathfrak{c}_{t,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t,a_t\right]\\ +\mathfrak{c}_{s,T}\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\frac{K_h(x_t-x_s)\mathbf{1}[a_t=a_s]}{f(x_s,a_s)}-\mathfrak{c}_{s,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_s,a_s\right]\end{array}\right) \\
& -\sqrt{T}\left(\begin{array}{c}T-1\\2\end{array}\right)^{-1}\sum_{t=1}^{T-1}\sum_{s>t}\frac{1}{2}\left(\begin{array}{c}\mathfrak{c}_{t,T}\frac{K_h(x_s-x_t)\mathbf{1}[a_s=a_t]-f_{X,A}(x_t,a_t)}{f_{X,A}(x_t,a_t)}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t,a_t\right]\\ +\mathfrak{c}_{s,T}\frac{K_h(x_t-x_s)\mathbf{1}[a_t=a_s]-f(x_s,a_s)}{f(x_s,a_s)}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_s,a_s\right]\end{array}\right)+o_p(1).
\end{aligned}
$$

Hoeffding (H-)decomposition provides the following as leading term, disposing the trimming factor,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}-E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t,a_t\right]\right). \tag{68}$$

To obtain the projection of the second term is more labor intensive. We first split it up into two parts,

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathfrak{c}_{t,T}\left[\mathcal{H}(I-\mathcal{L})^{-1}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}(x_t,a_t)\right] \\
= & \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathfrak{c}_{t,T}\left[\mathcal{H}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}(x_t,a_t)\right]+\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathfrak{c}_{t,T}\left[\mathcal{H}\mathcal{L}(I-\mathcal{L})^{-1}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\frac{\partial^k m_{\theta_0}}{\partial\theta_j^k}(x_t,a_t)\right].
\end{aligned}
$$

The summands of the first term takes the following form

$$\mathfrak{c}_{t,T}\beta\int\left(\begin{array}{c}\int\frac{\partial m_{\theta_0}(x'')}{\partial\theta_j}\frac{1}{f_X(x')}\left(\widehat{f}_{X'X}(dx'',x')-f_{X'X}(dx'',x')\right)\\ -\frac{\widehat{f}_X(x')-f_X(x')}{f_X(x')}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+2})}{\partial\theta_j}\right|x_{t+1}=x'\right]\end{array}\right)f_{X'|X,A}(dx'|x_t,a_t),$$

with standard change of variable and usual symmetrization, this leads to the following kernel for the U-statistic,

$$
\begin{aligned}
& \frac{\beta}{2}\left(\begin{array}{c}\mathfrak{c}_{t,T}\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\frac{f'_{X|X,A}(x_s|x_t,a_t)}{f_X(x_s)}-\mathfrak{c}_{t,T}E\left[\left.E\left[\left.\frac{\partial m_{\theta_0}(x_{t+2})}{\partial\theta_j}\right|x_{t+1}\right]\right|x_t,a_t\right]\\ +\mathfrak{c}_{s,T}\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\frac{f'_{X|X,A}(x_t|x_s,a_s)}{f_X(x_t)}-\mathfrak{c}_{s,T}E\left[\left.E\left[\left.\frac{\partial m_{\theta_0}(x_{s+2})}{\partial\theta_j}\right|x_{s+1}\right]\right|x_s,a_s\right]\end{array}\right) \\
& -\frac{\beta}{2}\left(\begin{array}{c}\mathfrak{c}_{t,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\right|x_s\right]\frac{f'_{X|X,A}(x_s|x_t,a_t)}{f_X(x_s)}-\mathfrak{c}_{t,T}E\left[\left.E\left[\left.\frac{\partial m_{\theta_0}(x_{t+2})}{\partial\theta_j}\right|x_{t+1}\right]\right|x_t,a_t\right]\\ +\mathfrak{c}_{s,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t\right]\frac{f'_{X|X,A}(x_t|x_s,a_s)}{f_X(x_t)}-\mathfrak{c}_{s,T}E\left[\left.E\left[\left.\frac{\partial m_{\theta_0}(x_{s+2})}{\partial\theta_j}\right|x_{s+1}\right]\right|x_s,a_s\right]\end{array}\right),
\end{aligned}
$$

The leading term from H-decomposition leads to the following centered process

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\beta\left(\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}-E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t\right]\right), \tag{69}$$

notice the conditional expectation term is a two-step ahead predictor, zero mean follows from stationarity assumption and the law of iterated expectation. As for the second part of the second term, using the Neumann series representation, see (50) and (51), the kernel of the relevant U-statistics is,

$$\frac{\beta}{2}\left(\begin{array}{c} \mathfrak{c}_{t,T}\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\int\frac{\varphi(x_s|x')}{f_X(x_s)}f'_{X|X,A}\left(dx'|x_t,a_t\right)-\mathfrak{c}_{t,T}\sum_{\tau=1}^{\infty}\beta^{\tau}E\left[E\left[\left.\frac{\partial m_{\theta_0}(x_{t+\tau+2})}{\partial\theta_j}\right|x_{t+1}\right]\middle|x_t,a_t\right] \\ +\mathfrak{c}_{s,T}\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\int\frac{\varphi(x_t|x')}{f_X(x_t)}f'_{X|X,A}\left(dx'|x_s,a_s\right)-\mathfrak{c}_{s,T}\sum_{\tau=1}^{\infty}\beta^{\tau}E\left[E\left[\left.\frac{\partial m_{\theta_0}(x_{s+\tau+2})}{\partial\theta_j}\right|x_{s+1}\right]\middle|x_s,a_s\right] \end{array}\right)$$

$$-\frac{\beta}{2}\left(\begin{array}{c} \mathfrak{c}_{t,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{s+1})}{\partial\theta_j}\right|x_s\right]\int\frac{\varphi(x_s|x')}{f_X(x_s)}f'_{X|X,A}\left(dx'|x_t,a_t\right)-\mathfrak{c}_{t,T}\sum_{\tau=1}^{\infty}\beta^{\tau}E\left[E\left[\left.\frac{\partial m_{\theta_0}(x_{t+\tau+2})}{\partial\theta_j}\right|x_{t+1}\right]\middle|x_t,a_t\right] \\ +\mathfrak{c}_{s,T}E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t\right]\int\frac{\varphi(x_t|x')}{f_X(x_t)}f'_{X|X,A}\left(dx'|x_s,a_s\right)-\mathfrak{c}_{s,T}\sum_{\tau=1}^{\infty}\beta^{\tau}E\left[E\left[\left.\frac{\partial m_{\theta_0}(x_{s+\tau+2})}{\partial\theta_j}\right|x_{s+1}\right]\middle|x_s,a_s\right] \end{array}\right.$$

where φ is defined as the limit of discounted sum of the conditional densities, see (52). The projection of the U-statistic with the above kernel yields,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\frac{\beta^2}{1-\beta}\right)\left(\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}-E\left[\left.\frac{\partial m_{\theta_0}(x_{t+1})}{\partial\theta_j}\right|x_t\right]\right). \tag{70}$$

The last term of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{1,t,T}\right)_j$ can be treated similarly, recall we have

$$\begin{aligned} &\mathcal{H}\left(I-\mathcal{L}\right)^{-1}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right) \\ =\;&\mathcal{H}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)+\mathcal{H}\mathcal{L}\left(I-\mathcal{L}\right)^{-1}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right). \end{aligned}$$

Ignoring the bias term, that is negligible under assumptions B6 and B7,

$$\begin{aligned} &\mathcal{H}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)(x_t,a_t) \\ =\;&\frac{1}{T-1}\sum_{\widetilde{a}\in A}\sum_{s\neq t}\int\left(\frac{\partial^k\zeta'_{x,\widetilde{a},\theta_0}\left(P\left(\widetilde{a}|x'\right)\right)}{\partial\theta_j^k}\frac{e_{\widetilde{a},s}K_h\left(x_s-x'\right)}{f_X\left(x'\right)}\right)f'_{X|X,A}\left(dx'|x_t,a_t\right)+o_p\left(T^{-1/2}\right), \\ =\;&\frac{1}{T-1}\sum_{\widetilde{a}\in A}\sum_{s\neq t}f'_{X|X,A}\left(x_s|x_t,a_t\right)\frac{\partial^k\zeta'_{x_s,\widetilde{a},\theta_0}\left(P\left(\widetilde{a}|x\right)\right)}{\partial\theta_j^k}\frac{e_{\widetilde{a},s}}{f_X\left(x_s\right)}+o_p\left(T^{-1/2}\right). \end{aligned}$$

Normalizing the projection of the corresponding U-statistics obtains

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{H}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)(x_t,a_t)=\frac{1}{\sqrt{T}}\sum_{\widetilde{a}\in A}\sum_{t=1}^{T}\frac{\partial^k\zeta'_{x_t,\widetilde{a},\theta_0}\left(P\left(\widetilde{a}|x_t\right)\right)}{\partial\theta_j^k}e_{\widetilde{a},t}+o_p\left(1\right). \tag{71}$$

The same can be done to the remaining term, in particular we obtain

$$\begin{aligned} &\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\mathcal{H}\mathcal{L}\left(I-\mathcal{L}\right)^{-1}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)(x_t,a_t) \\ =\;&\frac{1}{\sqrt{T}}\sum_{\widetilde{a}\in A}\sum_{t=1}^{T}\frac{\beta}{1-\beta}\frac{\partial^k\zeta'_{x_t,\widetilde{a},\theta_0}\left(P\left(\widetilde{a}|x_t\right)\right)}{\partial\theta_j^k}e_{\widetilde{a},t}+o_p\left(1\right). \end{aligned} \tag{72}$$

Collecting (68) - (72), for $k=1$,we obtain the leading terms of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{1,t,T}\right)_j$. For $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{2,t,T}\right)_j$ and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{3,t,T}\right)_j$, we again use the projection technique of the U-statistics to obtain their leading terms. We gave a lot of details for the former case as remaining terms in $(D_{2,T})$ can be treated in a similar fashion. In particular, it is simple to show that the projections of various relevant U-statistics, defined below with some elements $\varpi_k \in C(X), \varsigma_k \in C(A\times X)$ and $\widetilde{a}\in A$, have the following linear representation:

1. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varsigma_k(x_t,\widetilde{a})\left[\left(\widehat{\mathcal{H}}-\mathcal{H}\right)\varpi_k(x_t,\widetilde{a})\right]$
$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\left(\frac{\varsigma_k(x_t,\widetilde{a})f_X(x_t)\mathbf{1}[a_t=\widetilde{a}]}{f(x_t,\widetilde{a})}\right)\left(\varpi_k(x_{t+1})-E\left[\varpi_k(x_{t+1})\mid x_t,a_t=\widetilde{a}\right]\right)+o_p(1)$.
2. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varsigma_k(x_t,\widetilde{a})\left[\mathcal{H}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\varpi_k(x_t,\widetilde{a})\right]$
$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\beta\left(\frac{\int\varsigma_k(v,\widetilde{a})f'_{X|X,A}(x_t|v,\widetilde{a})f_X(dv)}{f_X(x_t)}\right)\left(\varpi_k(x_{t+1})-E\left[\varpi_k(x_{t+1})\mid x_t\right]\right)+o_p(1)$.
3. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varsigma_k(x_t,\widetilde{a})\left[\mathcal{H}\mathcal{L}(I-\mathcal{L})^{-1}\left(\widehat{\mathcal{L}}-\mathcal{L}\right)\varpi_k(x_t,\widetilde{a})\right]$
$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\beta\left(\frac{\int\int\varsigma_k(v,\widetilde{a})\varphi(x_t|w)f'_{X|X,A}(dw|dv,\widetilde{a})f_X(dv)}{f_X(x_t)}\right)\left(\varpi_k(x_{t+1})-E\left[\varpi_k(x_{t+1})\mid x_t\right]\right)+o_p(1)$.

In correspondence of $\left(E_{k+1,t,T}\right)_j$ for $k=1,2$, we have in mind

$$\begin{aligned}\varsigma_1(\cdot) &= \psi_1(\widetilde{a},\cdot),\\ \varpi_1(\cdot) &= \frac{\partial m_{\theta_0}(\cdot)}{\partial\theta_j},\\ &\text{and}\\ \varsigma_2(\cdot) &= \psi_{2,j}(\widetilde{a},\cdot),\\ \varpi_2(\cdot) &= m_{\theta_0}(\cdot),\end{aligned}$$

where ψ_1 and $\psi_{2,j}$ are defined in (66) - (67). Similarly, we also have

4. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varsigma_k(x_t,\widetilde{a})\mathcal{H}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)(x_t,\widetilde{a})$
$= \frac{1}{\sqrt{T}}\sum_{a^*\in A}\sum_{t=1}^{T-1}\left(\int\varsigma_k(v,\widetilde{a})f'_{X|X,A}(x_t|v,\widetilde{a})f_X(dv)\right)\frac{\partial^k\zeta'_{x_t,\widetilde{a},\theta_0}(P(a^*|x_t))}{\partial\theta_j^k}\frac{e_{a^*,t}}{f_X(x_t)}+o_p(1)$.
5. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varsigma_k(x_t,\widetilde{a})\left[\mathcal{H}\mathcal{L}(I-\mathcal{L})^{-1}\left(\frac{\partial^k\widehat{r}_{\theta_0}}{\partial\theta_j^k}-\frac{\partial^k r_{\theta_0}}{\partial\theta_j^k}\right)(x_t,\widetilde{a})\right]$
$= \frac{1}{\sqrt{T}}\sum_{a^*\in A}\sum_{t=1}^{T}\left[\int\int\varsigma_k(v,\widetilde{a})\varphi(x_t|w)f'_{X|X,A}(dw|v,\widetilde{a})f_X(dv)\right]\frac{\partial^k\zeta'_{x_t,\widetilde{a},\theta_0}(P(a^*|x_t))}{\partial\theta_j^k}\frac{e_{a^*,t}}{f_X(x_t)}+o_p(1)$.

Notice that leading terms from all the projections above are mean zero processes. Collecting these terms, lots of covariance. Clearly $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{k,t,T}\right)_j=O_p(1)$ for $k=1,2,3$ and $j=1,\ldots,q$, this ensures the root-T consistency $\widehat{\theta}$. The term $D_{3,T}$ is $o_p(1)$ since $\frac{\partial q\left(a_t,x_t;\theta_0,g_{\theta_0}\right)}{\partial\theta}$ is uniformly bounded and $\mathfrak{d}_{t,T}=o_p\left(\sqrt{T}\right)$ for all t. In sum,

$$\sqrt{T}D_{2,T}\Rightarrow N(0,\Lambda_2),$$

$$\Lambda_2=\lim_{T\to\infty}Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(E_{1,t,T}+E_{2,t,T}+E_{3,t,T}\right)\right),$$

$$\sqrt{T}\left(\widehat{\theta}-\theta_0\right) \Rightarrow N\left(0, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right),$$

$$\begin{aligned}
\mathcal{I} &= \lim_{T\to\infty} Var\left(D_{1,T}+D_{2,T}\right), \\
\mathcal{J} &= E\left[\frac{\partial^2 q\left(a_t, x_t; \theta_0, g_{\theta_0}\right)}{\partial\theta\partial\theta^\top}\right].
\end{aligned}$$

■

PROOF OF THEOREM 4 AND 5: Under the assumed smoothness assumptions, the results simply follow from MVT. ■

Appendix B

We now show various centered processes in the previous section converge uniformly at desired rates on a compact set X. We outline the main steps below and proof the results for relevant cases. The methodology here is similar to [LM] who employed the exponential inequality from [B] for various quantities similar to ours.

Consider some process $\mathsf{l}_T(x) = \frac{1}{T}\sum l(x_t, x)$, where $l(x_t, x)$ has mean zero. For some positive sequence, δ_T, converging monotonically to zero, we first show that $|\mathsf{l}_T(x)| = o_p(\delta_T)$ pointwise on X, then we use the continuity property of $l(x_t, x)$ to show that this rate of convergence is preserved uniformly over X.

To obtain the pointwise rates, specializing Theorem 1.3 of [B], we have the following inequality.

$$\begin{aligned}
\Pr\left(|\mathsf{l}_T(x)| > \delta_T\right) &\leq 4\exp\left(-\frac{\delta_T^2 T^\beta}{8v^2\left(T^\beta\right)}\right) + 22\left(1+\frac{4b_T}{\delta_T}\right)^{1/2} T^\beta \alpha\left(\left\lfloor \frac{T^{1-\beta}}{2}\right\rfloor\right) \\
&\leq \exp\left(-G_{1,T}\right) + G_{2,T},
\end{aligned}$$

for some

$$\begin{aligned}
\beta &\in (0,1), \\
b_T &= O\left(\sup_{(x',x)\in\mathcal{X}\times\mathcal{X}} l(x',x)\right), \\
v^2(\beta) &= \text{var}\left(\frac{1}{\left\lfloor \frac{T^\beta}{2}+1\right\rfloor}\sum_{t=1}^{\left\lfloor \frac{T^\beta}{2}+1\right\rfloor} l(x_t,x)\right) + \frac{b_T\delta_T}{2}.
\end{aligned} \tag{73}$$

To have the first term converging to zero, at an exponential rate, we need $G_{1,T} \to \infty$. The main calculation here is the variance term in v^2. Following [M], we can generally show that the uniform order of such term comes from the variances and the covariances terms are of smaller order. We note that the bounds on these variances are independent on the trimming set. For our purposes, the natural choice of δ_T^2 often reduces us to choosing β to satisfy $b\delta_T = o\left(\delta_T^2 T^{\beta}\right)$. The rate of $G_{2,T}$ is easy to control since all of the quantities involved increase (decrease) at a power rate, the mixing coefficient can be made to decay sufficiently fast so $G_{2,T} = O\left(T^{-\eta}\right)$ for some $\eta > 0$, hence $\Pr\left(\left|\mathsf{l}_T(x)\right| > \delta_T\right) = O\left(T^{-\eta}\right)$.

To obtain the uniform rates over $\mathcal{X}$, compactness implies there exist an increasing number, Q_T, of shrinking hyper-cubes $\{I_{Q_T}\}$ whose length of each side is $\{\epsilon_T\}$ with centres $\left\{x^{Q_T}\right\}$. These cubes cover $\mathcal{X}$, namely for some C_0 and d, $\epsilon_T^d Q_T \leq C_0 < \infty$. In particular, we will have Q_T grow at a power rate in our applications. Then we have

$$\begin{aligned}
\Pr\left(\sup_x \left|\mathsf{l}_T(x)\right| > \delta_T\right) &\leq \Pr\left(\max_{1\leq q\leq Q_T}\left|\mathsf{l}_T(x^q)\right| > \delta_T\right) + \Pr\left(\max_{1\leq q\leq Q_T}\sup_{x\in I_q}\left|\mathsf{l}_T(x) - \mathsf{l}_T(x^q)\right| > \delta_T\right) \\
&= G_{3,T} + G_{4,T},
\end{aligned}$$

where $G_{3,T} = O\left(Q_T T^{-\eta}\right)$ by Bonferroni Inequality. Provided the rate of decay of the mixing coefficient, i.e. η, is sufficiently large relative to the rate Q_T grows we shall have $Q_T = o\left(T^{\eta}\right)$. For the second term, since the opposing behavior of (ϵ_T, Q_T) is independent of the mixing coefficient, $\max_{1\leq q\leq Q_T}\sup_{x\in I_q}\left|\mathsf{l}_T(x) - \mathsf{l}_T(x^q)\right| = o\left(\delta_T\right)$ can be shown using Lipschitz continuity when the hyper cubes shrink sufficiently fast.

Before we proceed with the specific cases we validate our treatment of the trimming factor. The pointwise rates are clearly unaffected by bias at the boundary so long $x \in \operatorname{int}(X)$. The technique used to obtain uniformity also accommodates expanding space $\mathcal{X}$, so long we use the sequence $\{c_T\}$ to satisfy condition stated in [B9]. The uniform rate of convergence is also unaffected, when replace $\mathcal{X}$ with X_T, since the covering of an expanding of a compact subsets of a compact set can still grow (and shrink) at the same rate in each of the cases below. Therefore we could replace $\mathcal{X}$ everywhere by X_T.

Combining the results of uniform convergence of the zero mean processes and their biases, the uniform rates to various quantities in the previous section can now be established. We note that the treatment to allow for additional discrete observable states only requires trivial extension. We provide illustrate this for the first case of kernel density estimation, and for brevity, thenceforth assume that we only have purely continuous observable state variables.

Case 1. Here we deal with density estimators such as $\widehat{f}_X(x), \widehat{f}_{X',X}(x',x)$ and $\widehat{f}_{X',X,A}(x',x,j)$:

We first establish the pointwise rate of convergence of a de-meaned kernel density estimator.

$$\begin{aligned}
\mathsf{l}_T(x) &= \widehat{f}_X(x) - E\widehat{f}_X(x), \\
l(x_t, x) &= \prod_{l=0}^{d-1} K_h\left(x_{t-l} - x_{l+1}\right) - E\prod_{l=0}^{d-1} K_h\left(x_{t-l} - x_{l+1}\right).
\end{aligned}$$

The main elements for studying the rate of $G_{1,T}$ are

$$\begin{aligned}
\varpi &= \frac{1}{\sqrt{Th^d}}, \\
\delta_T &= T^{\xi}\varpi \quad \text{for some} \quad \xi > 0, \\
b_T &= O\left(h^{-d}\right), \\
v^2\left(T^{\beta}\right) &= O\left(\varpi^2 T^{1-\beta} \vee T^{\xi}\varpi h^{-d}\right).
\end{aligned}$$

We obtain from simple algebra

$$G_{1,T} = O\left(\frac{T^{2\xi}T^{\beta}}{T^{1-\beta} + T^{\xi}T^{1/2}h^{-d/2}}\right).$$

As mentioned in the previous section, we have $d = 2$ and $h = O\left(T^{-1/5}\right)$. This means $\delta_T = T^{\xi-3/10}$, and if $\beta \in (7/10, 1)$ then we have $G_{1,T} \to \infty$. Clearly, the same choice of β will suffice for $d = 1$ as well.

To make this uniform on X_T, with product kernels and the Lipschitz continuity of K, we have for any $(x, x_q) \in I_q$,

$$\left|K_h\left(x_t - x\right) - K_h\left(x_t - x_q\right)\right| \leq \frac{C_1}{h^3}\epsilon_T.$$

So it follows that

$$\begin{aligned}
\delta_T^{-1} \max_{1\leq q\leq Q_T} \sup_{x\in I_q} \left|\mathsf{l}_T(x) - \mathsf{l}_T(x_q)\right| &= O\left(\frac{\epsilon_T}{\delta_T h^3}\right) \\
&= O\left(\frac{T^{-\zeta/2}}{T^{\xi-9/10}}\right).
\end{aligned}$$

Define $Q_T = T^{\zeta}$, for some $\zeta > 0$, this requires $9/5 < \zeta < \eta$.

We can allow for additional discrete control variable and/or observable state variables. As an illustration, consider the density estimator of one continuous random variable and some discrete random variable, we have

$$\begin{aligned}
\mathsf{l}_T(x) &= \widehat{f}_{X^C,X^D}\left(x_c, x_d\right) - E\widehat{f}_{X^C,X^D}\left(x_c, x_d\right), \\
l\left(x_t, x\right) &= K_h\left(x_{c,t} - x_c\right)\mathbf{1}\left(x_{d,t} = x_d\right) - EK_h\left(x_{c,t} - x_c\right)\mathbf{1}\left(x_{d,t} = x_d\right).
\end{aligned}$$

Same rates as the purely continuous case apply. For the pointwise part, the variance is clearly of the same order. For the bounds on the uniform rates observe that,

$$\left|K_h\left(x_{c,t} - x_c\right)\mathbf{1}\left(x_{d,t} = x_d\right) - K_h\left(x_{c,t} - x_c^q\right)\mathbf{1}\left(x_{d,t} = x_d\right)\right| \leq \left|K_h\left(x_{c,t} - x_c\right) - K_h\left(x_{c,t} - x_c^q\right)\right|.$$

Same reasoning also applies for the kernel estimator of the density of the control and observable state variables.

CASE 2. Here we deal with $\widehat{r}^{C}(x)$:

$$l(x_t,x)=\frac{e_{\theta,t}K_h(x_t-x)}{f_X(x)}.$$

Since $\{e_{\theta,T}\}$ is uniformly bounded (a.s.) it follows, as shown in Case 1, the choice $\beta\in(3/5,1)$ will do to have $G_{1,T}\to\infty$.

To make this uniform on X_T, by boundedness of $\{e_{\theta,T}\}$, Lipschitz continuity of K,f and their appropriate bounds, we have for any $(x,x_q)\in I_q$,

$$|K_h(x_t-x)-K_h(x_t-x_q)|\leq\frac{C_2}{h^2}\epsilon_T.$$

So it follows that

$$\begin{aligned}\delta_T^{-1}\max_{1\leq q\leq Q_T}\sup_{x\in I_q}|\mathsf{l}_T(x)-\mathsf{l}_T(x_q)| &= O\left(\frac{T^{-\zeta}}{T^{\xi-7/10}}\right)\\ &= o(1),\end{aligned}$$

for some $\zeta>0$, this requires $7/10<\zeta<\eta$.

CASE 3. Here we deal with $\mathcal{L}(I-\mathcal{L})^{-1}\widehat{r}_\theta^{C}(x)$:

$$l(x_t,x)=e_{\theta,t}\nu(x_t,x),$$

where the definition of ν is provided in (54). Using Billingsley's Inequality, it is straightforward to show that with the additional smoothing, the variance of l_T is of parametric rate uniformly on X_T. Selecting $\beta\in(1/2,1)$ will yield $G_{1,T}\to\infty$ for $\Pr\left(|\mathsf{l}_T(x)|>T^{-2/5}\right)=o(1)$, for any $x\in X_T$.

To make this uniform on X_T, by boundedness of $\{e_{\theta,T}\}$ and Lipschitz continuity of φ, we have for any $(x,x_q)\in I_q$,

$$|e_{\theta,t}\nu(x_t,x)-e_{\theta,t}\nu(x_t,x_q)|\leq C_3\epsilon_T.$$

So it follows that

$$\delta_T^{-1}\max_{1\leq q\leq Q_T}\sup_{x\in I_q}|\mathsf{l}_T(x)-\mathsf{l}_T(x_q)|=O\left(\frac{T^{-\zeta}}{T^{-2/5}}\right),$$

for some $\zeta>0$, this requires $2/5<\zeta<\eta$.

CASE 4. Here we deal with $m_{1,\theta}^{C}(x)$:

$$\mathsf{l}_T(x)=\frac{\beta}{f_X(x)}\int\left(\widehat{f}_{X',X}(x',x)-E\widehat{f}_{X',X}(x',x)\right)m_\theta(x')\,dx'.$$

As mentioned in the previous section, under our smoothness assumptions, we have uniformly on X_T,

$$\int\widehat{f}_{X',X}(x',x)\,m_\theta(x')\,dx'=\frac{1}{T-1}\sum_{t=1}^{T-1}K_h(X_t-x)\,m_\theta(X_{t+1})+O\left(h^2\right).$$

The exact same choices found in Case 2 apply. ■

References

[1] Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2005): "Econometric Tools for Analyzing Market Outcome," *Handbook of Econometrics*, vol. 6, eds. J. Heckman and E. Leamer. North-Holland.

[2] Aguirregabiria, V., and P. Mira (2002): "Swapping Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, **70**, 1519-1543.

[3] Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, **75**, 1-53.

[4] Aguirregabiria, V., and P. Mira (2008): "Dynamic Discrete Choice Structural Models: A Survey," *Journal of Econometrics,* forthcoming.

[5] Aitchison, J., and C. G. G. Aitken (1976): "Multivariate Binary Discrimination by the Kernel Method," *Biometrika,* **63**, 413-420.

[6] Bajari, P., C. L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition" *Econometrica*, **75**, 1331-1370.

[7] Bajari, P., V. Chernozhukov, H. Hong and D. Nekipelov (2008): "Semiparametric Estimation of a Dynamic Game of Incomplete Information," *Working paper.*

[8] Bellman, R. E. (1957): *Dynamic Programming,* Princeton University Press.

[9] Bertsekas, D. P.and S. Shreve (1978): *Stochastic Optimal Control: The Discrete-TIme Case*, Academic Press

[10] Bosq, D. (1996) (1996): *Nonparametric Statistics for Stochastic Processes*, Springer.

[11] Carrasco, M., J.P. Florens and E. Renault (2007): "Linear Inverse Problems and Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, vol. 6, eds. J. Heckman and E. Leamer. North-Holland.

[12] Chen, X., O.B. Linton and I. van Keilegom (2003): "Estimation of Semiparametric Models when the Criterion Function is not Smooth," *Econometrica*, **71**, 1591-1608.

[13] Fan, J. (1993): "Local Linear Regression Smoothers and their Minimax Efficiencies," *Annal of Statistics,* **21**, 196-216.

[14] Granas, A. and J. Dugundji (2003): *Fixed Point Theory*, Springer.

[15] Hajivassiliou, V., and P. Ruud (1994): "Classical Estimation Methods for LDV Models Using Simulation," *Handbook of Econometrics*, vol. 4, eds. R.F. Engle and D.L. McFadden. North-Holland.

[16] Hotz, V., and R.A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, **60**, 497-531.

[17] Hotz, V., R.A. Miller, S. Smith and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, **61**, 265-289.

[18] Jofre-Bonet, M., and M. Pesendorfer (2003): "Estimation of a Dynamic Auction Game," *Econometrica*, **71**, 1443-1489.

[19] Judd, K. (1998): *Numerical Methods in Economics,* MIT Press, Cambridge MA.

[20] Kress, R. (1999): *Linear Integral Equations*, Springer.

[21] Linton, O.B., and E. Mammen (2003): "Estimating Semiparametric Arch(∞) Models by Kernel Smoothing Methods," STICERD working paper EM/2003/453.

[22] Linton, O.B., and E. Mammen (2005): "Estimating Semiparametric Arch(∞) Models by Kernel Smoothing Methods," *Econometrica*, **73**, 771-836.

[23] Masry, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," Journal of Time Series Analysis, **17**, 571 - 599.

[24] McFadden, D.L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontier in Econometrics*, eds P. Zarembka. New York.

[25] Newey, W.K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica,* **62**, 1349-1382.

[26] Pakes, A., and S. Olley (1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics,* **65**, 295-332.

[27] Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica,* **57**, 1027-1057.

[28] Pakes, A., M. Ostrovsky, and S. Berry (2004): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Example)," *Working Paper.*

[29] Pesendorfer, M., and P. Schmidt-Dengler (2003): "Identification and Estimation of Dynamic Games," *Working Paper.*

[30] Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimator for Dynamic Games," *Reviews of Economics Studies*, **75**, 901-928.

[31] Powell, J.L., J. H. Stock and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica,* **57**, 1403-1430.

[32] Robinson, P.M. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis,* **4**, 185-207.

[33] Robinson, P.M. (1988): "Root-N-consistent semiparametric regression," *Econometrica,* **56**, 931-954.

[34] Rosenblatt, M. (1956): "Remarks on some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics,* **27**, 832 - 837.

[35] Roussas, G. (1967): "Nonparametric Estimation in Markov Processes," *Annals of the Institute of Statistical Mathematics,* **21**, 73-87.

[36] Roussas, G. (1969): "Nonparametric Estimation of the Transition Distribution Function of a Markov Process," *Annals of Mathematical Statistics,* **40**, 1386-1400.

[37] Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica,* **55**, 999 - 1033.

[38] Rust, J. (1988): "Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal on Control and Optimization*, **26**, 1006-1023.

[39] Rust, J. (1994a): "Estimation of Dynamic Structural Models: Problems and Prospects Part I: Discrete Decision Processes," *Proceedings of the 6th World Congress of the Econometric Society, Cambridge University Press.*

[40] Rust, J. (1994b): "Structural Estimation of Markov Decision Processes," *Handbook of Econometrics*, vol. 4, eds. R.F. Engle and D.L. McFadden. North-Holland.

[41] Stokey, N.L., and R.E. Lucas (1989): *Recursive Methods in Economics Dynamics,* Harvard University Press, Cambridge MA.

T	bandwidth	bias	mbias	std	iqr	mse
100	$h^{1/2}$	0.0591	0.0584	0.4151	0.4244	0.1758
	h	0.0385	0.0351	0.4079	0.4140	0.1678
	$h^{3/2}$	0.0239	0.0216	0.4029	0.4020	0.1629
500	$h^{1/2}$	0.0530	0.0492	0.1752	0.1713	0.0335
	h	0.0306	0.0257	0.1712	0.1665	0.0303
	$h^{3/2}$	0.0193	0.0143	0.1694	0.1643	0.0291
1000	$h^{1/2}$	0.0408	0.0432	0.1249	0.1248	0.0173
	h	0.0184	0.0203	0.1224	0.1208	0.0153
	$h^{3/2}$	0.0082	0.0098	0.1212	0.1189	0.0147

Table 1: Summary statistics for $\widehat{\theta}_1$. $h = 1.06s(T)^{-1/5}$ is the bandwidth used in the nonparametric estimation, $s =$ denotes the standard deviation of $\{x_t\}_{t=1}^T$.

T	bandwidth	bias	mbias	std	iqr	mse
100	$h^{1/2}$	0.0806	0.0543	0.4806	0.4576	0.2374
	h	0.0606	0.0338	0.4913	0.4626	0.2450
	$h^{3/2}$	0.0479	0.0193	0.4974	0.4658	0.2497
500	$h^{1/2}$	0.0619	0.0544	0.2033	0.1993	0.0452
	h	0.0394	0.0328	0.2078	0.2062	0.0447
	$h^{3/2}$	0.0290	0.0224	0.2104	0.2085	0.0451
1000	$h^{1/2}$	0.0434	0.0375	0.1443	0.1444	0.0227
	h	0.0205	0.0150	0.1472	0.1512	0.0221
	$h^{3/2}$	0.0107	0.0048	0.1484	0.1527	0.0221

Table 2: Summary statistics for $\widehat{\theta}_2$. $h = 1.06s(T)^{-1/5}$ is the bandwidth used in the nonparametric estimation, $s =$ denotes the standard deviation of $\{x_t\}_{t=1}^T$.

		Values of x								
NT	bandwidth	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
100	$h^{1/2}$	0.0192	0.0134	0.0088	0.0050	0.0013	-0.0027	-0.0077	-0.0139	-0.0213
	(iqr)	0.0206	0.0151	0.0119	0.0101	0.0092	0.0104	0.0114	0.0146	0.0174
	h	0.0067	0.0039	0.0031	0.0027	0.0017	0.0003	-0.0016	-0.0047	-0.0098
	(iqr)	0.0341	0.0254	0.0226	0.0211	0.0213	0.0218	0.0202	0.0227	0.0264
	$h^{3/2}$	-0.0027	-0.0018	0.0003	0.0019	0.0022	0.0019	0.0016	0.0006	-0.0017
	(iqr)	0.0466	0.0381	0.0363	0.0331	0.0343	0.0337	0.0330	0.0328	0.0350
500	$h^{1/2}$	0.0151	0.0096	0.0058	0.0029	0.0005	-0.0020	-0.0052	-0.0097	-0.0159
	(iqr)	0.0117	0.0096	0.0077	0.0060	0.0058	0.0067	0.0074	0.0089	0.0107
	h	0.0027	0.0006	0.0002	0.0001	-0.0001	-0.0003	-0.0003	-0.0003	-0.0025
	(iqr)	0.0184	0.0160	0.0145	0.0130	0.0119	0.0136	0.0141	0.0148	0.0163
	$h^{3/2}$	-0.0039	-0.0029	-0.0022	-0.0009	-0.0003	0.0006	0.0015	0.0034	0.0043
	(iqr)	0.0253	0.0227	0.0214	0.0213	0.0193	0.0220	0.0209	0.0205	0.0213
1000	$h^{1/2}$	0.0131	0.0082	0.0049	0.0026	0.0006	-0.0015	-0.0043	-0.0084	-0.0142
	(iqr)	0.0092	0.0075	0.0060	0.0049	0.0045	0.0051	0.0060	0.0076	0.0091
	h	0.0006	-0.0005	-0.0005	0.0000	0.0004	0.0004	0.0004	0.0005	-0.0009
	(iqr)	0.0142	0.0120	0.0116	0.0102	0.0099	0.0106	0.0107	0.0120	0.0138
	$h^{3/2}$	-0.0050	-0.0037	-0.0027	-0.0009	0.0004	0.0012	0.0022	0.0036	0.0045
	(iqr)	0.0196	0.0174	0.0177	0.0175	0.0156	0.0167	0.0160	0.0178	0.0180

Table 3: Estimated mean deviation from the choice specific expected value $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t = x, a_t = 1\right]$ for various values of $x \in [0,1]$ and its corresponding interquartile range. $h = 1.06s(T)^{-1/5}$ is the bandwidth used in the nonparametric estimation, $s =$ denotes the standard deviation of $\{x_t\}_{t=1}^{T}$.

		Values of x								
NT	bandwidth	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
100	$h^{1/2}$	-0.0195	-0.0130	-0.0077	-0.0035	-0.0000	0.0035	0.0076	0.0129	0.0195
	(iqr)	0.0180	0.0144	0.0118	0.0111	0.0108	0.0108	0.0121	0.0148	0.0183
	h	-0.0075	-0.0041	-0.0025	-0.0014	-0.0004	0.0004	0.0016	0.0042	0.0086
	(iqr)	0.0285	0.0234	0.0218	0.0228	0.0248	0.0226	0.0221	0.0239	0.0297
	$h^{3/2}$	0.0011	0.0003	-0.0003	-0.0006	-0.0005	-0.0009	-0.0017	-0.0004	0.0010
	(iqr)	0.0374	0.0338	0.0359	0.0379	0.0399	0.0363	0.0332	0.0366	0.0422
500	$h^{1/2}$	-0.0157	-0.0095	-0.0049	-0.0018	0.0006	0.0029	0.0055	0.0091	0.0144
	(iqr)	0.0101	0.0085	0.0073	0.0068	0.0061	0.0065	0.0071	0.0089	0.0112
	h	-0.0029	-0.0000	0.0006	0.0003	0.0004	0.0007	0.0002	-0.0002	0.0017
	(iqr)	0.0152	0.0132	0.0142	0.0136	0.0135	0.0145	0.0138	0.0145	0.0170
	$h^{3/2}$	0.0032	0.0038	0.0031	0.0011	0.0001	0.0001	-0.0022	-0.0040	-0.0047
	(iqr)	0.0221	0.0206	0.0235	0.0225	0.0230	0.0238	0.0242	0.0232	0.0254
1000	$h^{1/2}$	-0.0138	-0.0082	-0.0043	-0.0016	0.0004	0.0024	0.0047	0.0080	0.0129
	(iqr)	0.0088	0.0076	0.0061	0.0052	0.0051	0.0057	0.0065	0.0077	0.0093
	h	-0.0005	0.0007	0.0004	0.0004	0.0002	-0.0000	-0.0004	-0.0006	0.0003
	(iqr)	0.0135	0.0124	0.0123	0.0115	0.0113	0.0119	0.0129	0.0127	0.0140
	$h^{3/2}$	0.0052	0.0040	0.0021	0.0016	0.0001	-0.0010	-0.0025	-0.0035	-0.0055
	(iqr)	0.0179	0.0180	0.0195	0.0190	0.0186	0.0191	0.0204	0.0199	0.0188

Table 4: Estimated mean deviation from the choice specific expected value $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t = x, a_t = 0\right]$ for various values of $x \in [0,1]$ and its corresponding interquartile range. $h = 1.06s(T)^{-1/5}$ is the bandwidth used in the nonparametric estimation, $s =$ denotes the standard deviation of $\{x_t\}_{t=1}^{T}$.

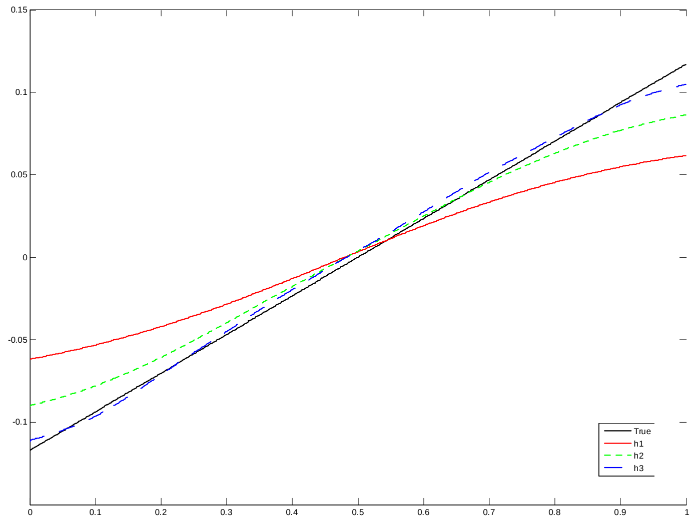

Figure 1: The difference in the estimated mean estimator of choice specific expected continuation values, $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 1\right] - E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 0\right]$, for $T = 100$; (h1,h2,h3) corresponds to the the bandwidths $(h^{1/2}, h, h^{3/2})$.

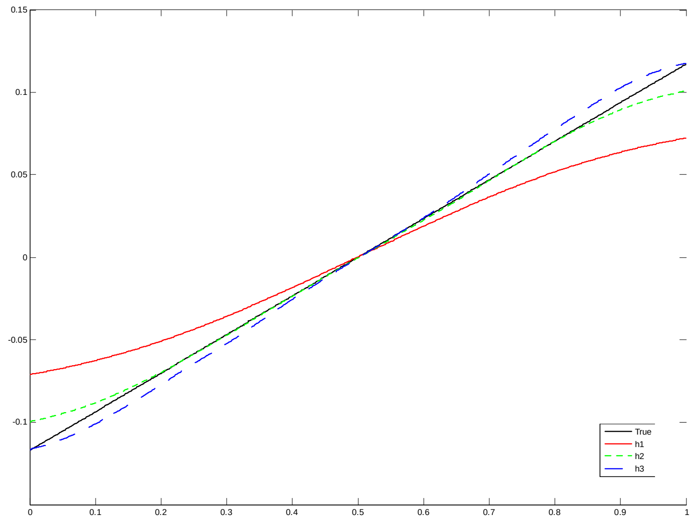

Figure 2: The difference in the estimated mean estimator of choice specific expected continuation values, $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 1\right] - E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 0\right]$, for $T = 500$; (h1,h2,h3) corresponds to the the bandwidths $(h^{1/2}, h, h^{3/2})$.

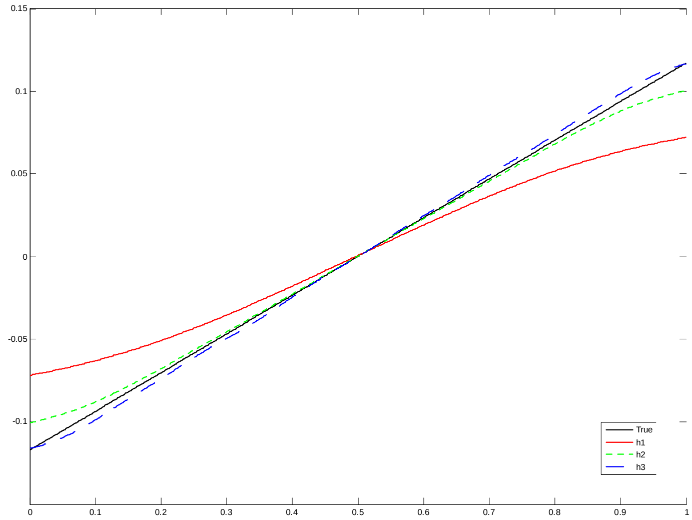

Figure 3: The difference in the estimated mean estimator of choice specific expected continuation values, $E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 1\right] - E\left[V_{\theta_0}\left(s_{t+1}\right)|x_t, a_t = 0\right]$, for $T = 1000$; (h1,h2,h3) corresponds to the the bandwidths $(h^{1/2}, h, h^{3/2})$.