# ON EXISTENCE OF BERK-NASH EQUILIBRIA IN MISSPECIFIED MARKOV DECISION PROCESSES WITH INFINITE SPACES

ROBERT M. ANDERSON, HAOSUI DUANMU, ANIRUDDHA GHOSH, AND M. ALI KHAN

Abstract. We present theorems on the existence of Berk-Nash equilibria in misspecified Markov Decision Processes with infinite action and state spaces. We extend the results of Esponda-Pouzo (2021) for finite state and action spaces to compact action spaces and sigma-compact state spaces with possibly unbounded payoff functions. This extension allows, for the first time, consideration of continuous distributions with possibly unbounded support. We provide several examples that span various areas in economic theory: neo-classical producer theory, the optimal savings problem, and identification and inference in econometric theory. The proofs use a recent technique in nonstandard analysis, originated by the second author, to extend known theorems for finite mathematical systems to infinite systems. This technique has already generated new results in probability theory, statistical decision theory, and general equilibrium theory, and is potentially applicable to a wide range of problems. (138 words)

## Contents



*Date*: 06/16/2022.

The authors are grateful to Demian Pouzo for helpful comments. Anderson gratefully acknowledges financial support from Swiss Re through the Consortium for Data Analytics in Risk.

## 1. Introduction

Learning under model misspecification is an important topic in economic theory.[1] Esponda and Pouzo (2021) (hereafter "EP") define an equilibrium notion, Berk-Nash Equilibrium, for misspecified Subjective Markov Decision Processes (SMDPs) in a dynamic programming environment.[2] EP presents results on the existence and stability of Berk-Nash equilibrium in SMDPs with *finite* state and action spaces. This paper extends the EP existence results to SMDPs with sigma-compact state and compact action spaces. This extension covers SMDPs with continuous distributions, unbounded state spaces and unbounded payoff functions. It covers examples chosen by EP to illustrate the importance of the Berk-Nash concept, but which lie outside the scope of their theorems.

The literature on Markov Decision Processes (MDPs) is replete with settings that naturally feature infinite state and action spaces.[3] We consider five examples from three important economic environments: (i) neoclassical producer theory, (ii) the optimal savings problem, and (iii) identification and inference in econometric theory. In the first environment, we consider two instances featuring demand and supply shocks to the revenues and the costs of the producer, and note the consequences of

[1]For literature on learning under misspecification, see Arrow and Green (1973), Nyarko (1991), Hansen and Sargent (2011), and Fudenberg et al. (2021) and the references therein.

[2]Classical references on dynamic programming include Maitra (1968), Puterman (1994), and Bhattacharya and Majumdar (2007).

[3]See Puterman (1994) for illustrations covering operations research, economics and engineering. Chapter 3 sketches many settings in economic theory that naturally demand infinite state and action spaces. Examples include the asset selling problem in Karlin (1962) and the employment seeking problem in Stokey and Lucas (1989).

the misspecified distributions of these shocks for the profit-maximizing choices.[4] The second environment extends Example 2 of EP, which features an optimal savings problem with a binary preference shock, to shocks with continuous and unbounded support.[5] Finally, we provide two examples in Gaussian AR(1) processes that connect the notion of Berk-Nash equilibrium to the existence of unit roots.[6]

In Theorem 1, we establish the existence of a Berk-Nash equilibrium for a *regular* misspecified SMDP with compact action, state and parameter spaces, and hence bounded payoff functions, but with unbounded densities (Radon-Nikodym derivatives). However, the assumption of a compact state spaces rules out distributions that have unbounded support, including normal, exponential and log-normal distributions which play central roles in economic theory and finance. The extension of Theorem 1 to more general environments necessitates further assumptions on the primitives of the problem. Theorem 2 considers a $\sigma$-compact state space with bounded payoff functions, under a regularity condition on the state space, a tightness condition on the class of probability measures and either a uniform integrability or a uniqueness condition on the relative entropy (Kullback-Liebler divergence). The tightness condition that we impose is satisfied by many economic applications (e.g. Ornstein-Uhlenbeck and Cox-Ingersoll-Ross processes) and we provide two sufficient conditions to test its applicability in environments of interest. Theorem 3 relaxes the boundedness[7] condition on the payoff functions by weaker bounded and continuity conditions (state-boundedness, fold-boundedness and W-continuity). Theorem 4 provides a possible learning foundation for SMDPs with compact state and action spaces that generalizes Theorem 2 in EP; since it depends on a strong convergence condition (see Definition C.6) that we hope to relax in future work, we report it in the Online Appendix.

The proofs make use of nonstandard analysis, a powerful mathematical technique that originated in Robinson (1966) and was introduced into mathematical economics in Brown and Robinson (1975). This paper makes use of a novel application developed recently by Duanmu (2018) and so far applied to probability theory, mathematical

[4]Lorenzoni (2009) explores the role of productivity shocks, news shocks and sampling shocks in driving business cycles with the shocks normally distributed with the real line as their support.

[5]A key paper that connects learning, optimal savings and uncertainty is Koulovatianos et al. (2009). Our convergence result (Theorem 4) in Appendix C applies potentially to such settings.

[6]See Farmer et al. (2021) for empirical illustrations connecting unit roots to model misspecification in macroeconomic settings. There are at least three more avenues where misspecification is being actively explored; climate economics (Berger and Marinacci (2020)), axiomatic decision theory (Cerreia-Vioglio et al. (2022)) and non-atomic anonymous games (Cerreia-Vioglio et al. (2020)).

[7]See Assumption 5 and Assumption 6 in Section 2.2 for further details.

statistics and economics.[8] Previous applications of nonstandard analysis in mathematical economics showed that results that are true in infinite settings but false in finite settings are approximately true in large finite settings.[9] Here, by contrast, we take results that are *exactly true* in finite settings and transport them to results in infinite settings. This method is applicable in situations in which the desired result is known for finite objects, its *proof* depends heavily on finiteness, but its *statement* makes sense for infinite objects.

A meta-theorem in mathematical logic guarantees that any nonstandard proof can be mechanically translated into a standard one, but the resulting standard proof can be very long and convoluted. The original EP proof can be modified to give a tractable standard proof for a version of Theorem 1 under the strong[10] additional assumption that the Radon-Nikodym derivatives are bounded. We do not know how to give tractable standard proofs for Theorem 1 without bounded Radon-Nikodym derivatives, for Theorem 2 or for Theorem 3.

Section 2 formalizes a misspecified SMDP, motivates the various assumptions we make on it, and presents our main results. Section 3 provides five concrete examples of SMDPs with infinite state and action spaces that are covered by our theorems. Section 4 lays out the methodological innovations of the paper and sketches the proofs. Section 5 briefly discusses the extensions of our single-agent results to a broader class of multi-agent misspecified environments. The Online Appendix gives a supplementary result that furnishes a learning foundation to the existence results and also contains the detailed analysis of the examples. The Appendix contains self-contained proofs of all our main results.

## 2. The Environment and Main Results

We begin by describing the environment faced by the agent which mirrors the one in EP. At the start of each period $t = 0, 1, 2, \dots$, the agent observes a state $s_t \in S$, takes an action $x_t \in X$ that determines the distribution of the future state $s_{t+1}$ given the transition probability function $Q(\cdot|s_t, x_t)$ with the initial state $s_0$,

[8] Duanmu's technique has previously been applied to statistical decision theory (Duanmu and Roy (2021)), Markov processes (Duanmu et al. (2021a) and Anderson et al. (2021b)), and to abstract economies and Walrasian equilibrium (Anderson et al. (2021a) and Anderson et al. (2022)).

[9] For previous applications of nonstandard analysis to mathematical economics, see for example, Brown and Robinson (1975), Anderson (1985), Khan (1976), Khan and Sun (2001), Duffie and Sun (2007), Anderson and Raimondo (2008), and Duffie et al. (2018).

[10] Given any two Gaussian distributions with distinct variances, the Radon-Nikodym derivative of the one with the larger variance with respect to the other is unbounded. Moreover, as we see in Example 3.1, unbounded Radon-Nikodym derivatives arise routinely in OLS estimation.

drawn according to the initial probability distribution $q_0$. For a given payoff function $\pi(s_t, x_t, x_{t+1})$, the agent then maximizes her expected discounted utility by choosing a feasible policy function. We now formally describe these objects.

**Definition 2.1.** A Markov Decision Process (MDP) is a tuple $\langle S, X, q_0, Q, \pi, \delta \rangle$, where

(1) The state space $S$ is a $\sigma$-compact locally compact metric space with Borel $\sigma$-algebra $\mathcal{B}[S]$;
(2) The action space $X$ is a compact metric space with Borel $\sigma$-algebra $\mathcal{B}[X]$;
(3) The initial distribution of states $q_0$ is a probability measure on $(S, \mathcal{B}[S])$;
(4) $Q : S \times X \to \Delta(S)$ is a transition probability function, where $\Delta(S)$ denotes the set of probability measures on $S$. That is, for each $(s, x) \in S \times X$, $Q(s, x)$ is a probability measure on $S$. We sometimes write $Q(\cdot|s, x)$ for $Q(s, x)(\cdot)$;
(5) $\pi : S \times X \times S \to \mathbb{R}$ is the per-period payoff function;
(6) The discount factor $\delta$ is in $[0, 1)$.

By the principle of optimality, the agent's problem can be cast recursively as

$$V(s) = \max_{x \in X} \int_S \{\pi(s, x, s') + \delta V(s')\} Q(\mathrm{d}s'|s, x). \tag{2.1}$$

Let $\mathcal{C}[S]$ denote the set of real-valued continuous functions equipped with the sup-norm. Then $\mathcal{C}[S]$ is a complete metric space. Let $F : \mathcal{C}[S] \to \mathcal{C}[S]$ be the operator such that $F(g)(s) = \max_{x \in X} \int_S \{\pi(s, x, s') + \delta g(s')\} Q(\mathrm{d}s'|s, x)$. Such $F$ is a contraction. By the Banach fixed point theorem, there exists an unique $V \in \mathcal{C}[S]$ that is a solution to the Bellman equation Eq. (2.1). We use MDP($Q$) to refer to Markov Decision Process with transition probability function $Q$.

**Definition 2.2.** An action $x$ is optimal given $s$ in the MDP($Q$) if

$$x \in \arg\max_{\hat{x} \in X} \int_S \{\pi(s, \hat{x}, s') + \delta V(s')\} Q(\mathrm{d}s'|s, \hat{x}) \tag{2.2}$$

We next describe a subjective Markov Decision Process.

**Definition 2.3.** A subjective Markov Decision Process is a Markov Decision Process, $\langle S, X, q_0, Q, \pi, \delta \rangle$, and a nonempty family of transition probability functions, $\mathcal{Q}_\Theta = \{Q_\theta : \theta \in \Theta\}$, where each transition probability function $Q_\theta : S \times X \to \Delta(S)$ is indexed by an element $\theta \in \Theta$. A subjective Markov Decision Process is said to be *misspecified* if $Q \notin \mathcal{Q}_\Theta$.

We write SMDP($\langle S, X, q_0, Q, \pi, \delta \rangle$, $\mathcal{Q}_\Theta$) to denote a subjective Markov Decision Process with the Markov Decision Process $\langle S, X, q_0, Q, \pi, \delta \rangle$ and the family $\mathcal{Q}_\Theta$ of transition probability functions. For all $\theta \in \Theta$, all $(s, x) \in S \times X$, let $D_\theta(\cdot|s, x) : S \to \bar{\mathbb{R}}$ be the density function if $Q(s, x)$ is dominated by $Q_\theta(s, x)$ and let $D_\theta(s'|s, x) = \infty$ if $Q(s, x)$ is not dominated by $Q_\theta(s, x)$.[11]

**Definition 2.4.** A regular subjective Markov decision process (regular-SMDP $\mathcal{M}$) is a SMDP that satisfies the following conditions:

(1) The parameter space $\Theta$ is a compact metric space;
(2) The mapping $(s, x) \to Q(s, x)$ is continuous in the Prokhorov metric;
(3) The mapping $(\theta, s, x) \to Q_\theta(s, x)$ is continuous in the Prokhorov metric;
(4) The density function $D_\theta(s'|s, x)$ is jointly continuous on the set $\{(\theta, s', s, x) : Q(s, x) \text{ is dominated by } Q_\theta(s, x)\}$;
(5) (Uniform integrability) For every compact set $S' \subset S$, there exists some $r > 0$ such that $\big(D_\theta(\cdot|s, x)\big)^{1+r}$ is uniformly integrable with respect to $Q_\theta(s, x)$ over the set $\{(\theta, s, x) : Q(s, x) \text{ is dominated by } Q_\theta(s, x)\}$. That is, for every $\epsilon > 0$, there exists $\kappa > 0$ such that

$$\int_E \big(D_{\theta_0}(t|s_0, x_0)\big)^{1+r} Q_{\theta_0}(s_0, x_0)(\mathrm{d}t) < \epsilon \tag{2.3}$$

if $(\theta_0, s_0, x_0)$ is an element of the set $\{(\theta, s, x) \in \Theta \times S' \times X : Q(s, x)$ is dominated by $Q_\theta(s, x)\}$ and $Q_{\theta_0}(s_0, x_0)(E) < \kappa$[12];
(6) (Absolute continuity) There is a dense set $\hat{\Theta} \subset \Theta$ such that $Q(s, x)$ is dominated by $Q_\theta(s, x)$ for all $\theta \in \hat{\Theta}$ and $(s, x) \in S \times X$;
(7) The per-period payoff function $\pi : S \times X \times S \to \mathbb{R}$ is continuous.

*Remark* 2.5. In Item 5, $\kappa$ depends on both $\epsilon$ and the compact set $S' \subset S$. Note that we allow $D_\theta(\cdot|s, x)$ to take value $\infty$ even if $Q(s, x)$ is dominated by $Q_\theta(s, x)$. So we allow for unbounded continuous density functions even when the state space is compact.

**Definition 2.6.** The weighted Kullback-Leibler divergence is a mapping $K_Q : \Delta(S \times X) \times \Theta \to \bar{\mathbb{R}}_{\geq 0}$ such that for any $m \in \Delta(S \times X)$ and $\theta \in \Theta$,

$$K_Q(m, \theta) = \int_{S \times X} \mathbb{E}_{Q(\cdot|s,x)} \left[\ln\big(D_\theta(s'|s, x)\big)\right] m(\mathrm{d}s, \mathrm{d}x). \tag{2.4}$$

[11] We use $\bar{\mathbb{R}}$ to denote the extended real line, equipped with the one-point compactification topology (Willard (2012)).
[12] This condition is automatically satisfied if the density functions $D_\theta(\cdot|s, x)$ are uniformly bounded over the set $\{(\theta, s, x) : Q(s, x) \text{ is dominated by } Q_\theta(s, x)\}$.

The set of closest parameter values given $m \in \Delta(S \times X)$ is the set[13]

$$\Theta_Q(m) = \arg\min_{\theta \in \Theta} K_Q(m, \theta). \tag{2.5}$$

For $(s, x) \in S \times X$ and $\theta \in \Theta$, the relative entropy (Kullback-Leibler divergence) from $Q_\theta(s, x)$ to $Q(s, x)$ is:

$$\mathcal{D}_{\mathrm{KL}}\big(Q(s, x), Q_\theta(s, x)\big) = \mathbb{E}_{Q(\cdot|s,x)}\left[\ln\big(D_\theta(s'|s, x)\big)\right]. \tag{2.6}$$

If $Q(s, x)$ is dominated by $Q_\theta(s, x)$, then we have

$$\mathcal{D}_{\mathrm{KL}}\big(Q(s, x), Q_\theta(s, x)\big) = \int_S D_\theta(s'|s, x) \ln\big(D_\theta(s'|s, x)\big) Q_\theta(\mathrm{d}s'|s, x). \tag{2.7}$$

and otherwise, it equals infinity. Moreover, by Item 5 in Definition 2.4, the function $D_\theta(\cdot|s, x) \ln\big(D_\theta(\cdot|s, x)\big)$ is an integrable function with respect to $Q_\theta(s, x)$. For $m \in \Delta(S \times X)$, let $\Theta_m = \{\theta \in \Theta : K_Q(m, \theta) < \infty\}$. By Definition 2.4, we have $\hat{\Theta} \subset \Theta_m$ and $K_Q(m, \theta)$ is a continuous function of $\theta$ on $\Theta_m$. Finally, by Jensen's inequality, the relative entropy $\mathcal{D}_{\mathrm{KL}}\big(Q(s, x), Q_\theta(s, x)\big)$ is non-negative for all $(s, x) \in S \times X$.

**Definition 2.7.** A probability distribution $m \in \Delta(S \times X)$ is a Berk-Nash equilibrium of the SMDP($\langle S, X, q_0, Q, \pi, \delta\rangle$, $\mathcal{Q}_\Theta$) if there exists a belief $\nu \in \Delta(\Theta)$ such that

(1) **Optimality**: For all $(s, x) \in S \times X$, that is in the support of $m$, $x$ is optimal given $s$ in the MDP($\bar{Q}_\nu$), where $\bar{Q}_\nu = \int_\Theta Q_\theta \nu(\mathrm{d}\theta)$;
(2) **Belief Restriction**: We have $\nu \in \Delta(\Theta_Q(m))$;
(3) **Stationarity**: For all $A \in \mathcal{B}[S]$, $m_S(A) = \int_{S \times X} Q(A|s, x) m(\mathrm{d}s, \mathrm{d}x)$, where $m_S$ denote the marginal measure of $m$ on $S$.

Next, we present three existence results: the first pertains to a compact state space and the other two, to non-compact state spaces. Throughout, we assume the action space $X$, and parameter space $\Theta$ to be compact.

2.1. **Existence of Equilibrium with Compact State Space.** Our first main result is on the existence of a Berk-Nash equilibrium when the underlying state space is compact.

**Theorem 1.** *For every regular-SMDP $\mathcal{M}$ with a compact state space, a Berk-Nash equilibrium exists.*

[13]We follow the standard convention in that $\ln(0) \cdot 0 = 0$ and integral of infinity over a set of measure 0 is 0. Further, $\frac{0}{0} = 0$, $\frac{1}{0} = \infty$, $\log \infty = \infty$.

2.2. **Existence of Equilibrium with Sigma-Compact State Space.** We extend Theorem 1 to a regular SMDP $\mathcal{M}$ with a non-compact state space. We assume that the density functions $\{D_\theta(\cdot|s,x)\}$ take value in $\mathbb{R}$. That is, for all $\theta \in \Theta$ and all $(s,x) \in S \times X$, let $D_\theta(\cdot|s,x) : S \to \mathbb{R}$ be the density function if $Q(s,x)$ is dominated by $Q_\theta(s,x)$ and let $D_\theta(s'|s,x) = \infty$ if $Q(s,x)$ is not dominated by $Q_\theta(s,x)$. We start with the following assumption on the state space $S$.

**Assumption 1.** (Regularity) There exists a non-decreasing sequence $\{S_n\}_{n\in\mathbb{N}}$ of compact subsets of $S$ such that

(1) $\bigcup_{n\in\mathbb{N}} S_n = S$;
(2) $q_0(S_n) > 0$ for all $n \in \mathbb{N}$;
(3) There exists $r > 0$ such that $Q(s,x)(S_n) > r$ and $Q_\theta(s,x)(S_n) > r$ for all $n \in \mathbb{N}$, all $(s,x) \in S_n \times X$ and all $\theta \in \Theta$;
(4) For all $n \in \mathbb{N}$, $S_n$ is a continuity set of $Q(s,x)$ and $Q_\theta(s,x)$ for all $(s,x) \in S_n \times X$ and all $\theta \in \Theta$.

Assumption 1 imposes four technical conditions on the state space that are satisfied for most applications in the literature. It requires that the state space $S$ can be deconstructed into a countable, non-decreasing sequence of subsets such that their union is the state space $S$. Items 3 and 4 of Assumption 1 jointly imply that the true and model transition probability functions are well-behaved for the truncation of the SMDP $\mathcal{M}$ defined on the sequence $\{S_n\}_{n\in\mathbb{N}}$. That is, for $n \in \mathbb{N}$, define $\mathcal{M}^n_{\Theta'} = (\langle S_n, X, q^n_0, Q^n, \pi_n, \delta\rangle,\ \mathcal{Q}^n_{\Theta'})$ to be the SMDP such that

(1) The state space is $S_n$, endowed with Borel $\sigma$-algebra $\mathcal{B}[S_n]$;
(2) The action space is $X$, endowed with Borel $\sigma$-algebra $\mathcal{B}[X]$;
(3) $q^n_0(A) = \frac{q_0(A)}{q_0(S_n)}$ for all $A \in \mathcal{B}[S_n]$;
(4) The parameter space $\Theta'$ is a finite subset of $\hat{\Theta}$;
(5) $Q^n : S_n \times X \to \Delta(S_n)$ is the transition probability function defined as $Q^n(s,x)(A) = \frac{Q(s,x)(A)}{Q(s,x)(S_n)}$ for all $A \in \mathcal{B}[S_n]$;
(6) For every $\theta \in \Theta'$, $Q^n_\theta : S_n \times X \to \Delta(S_n)$ is defined as $Q^n_\theta(s,x)(A) = \frac{Q_\theta(s,x)(A)}{Q_\theta(s,x)(S_n)}$ for all $A \in \mathcal{B}[S_n]$ and let $\mathcal{Q}^n_{\Theta'} = \{Q^n_\theta : \theta \in \Theta'\}$;
(7) $\pi_n : S_n \times X \times S_n \to \mathbb{R}$ is the restriction of $\pi$ to $S_n \times X \times S_n$;
(8) $\delta \in [0,1)$ is the discount factor.

*Remark* 2.8. For $\theta \in \Theta$, it is possible that $Q^n(s,x)$ is dominated by $Q^n_\theta(s,x)$ but $Q(s,x)$ is not dominated by $Q_\theta(s,x)$. Thus, we need to approximate the state and parameter spaces of the full SMDP $\mathcal{M}$ by carefully chosen subsets, simultaneously.

So we choose to approximate the state space $S$ by the sequence $\{S_n\}_{n\in\mathbb{N}}$ of compact sets and approximate the parameter space $\Theta$ by finite subsets of $\hat{\Theta}$.

To ensure that the Markov decision process has a stationary measure, a sequence of stationary measures for the truncated Markov decision processes should have a convergent subsequence. For every $n \in \mathbb{N}$ and every $P \in \Delta(S_n \times X)$, let $R_n(P)$ be the probability measure on $S_n$ such that

$$R_n(P)(A) = \int_{S_n \times X} Q^n(A|s,x)P(\mathrm{d}s, \mathrm{d}x). \tag{2.8}$$

Let $P_S$ denote the marginal measure of $P$ on $S$. The following tightness assumption guarantees the existence of a stationary measure.

**Assumption 2.** (Tightness) The family $\mathcal{R}$ is tight, where

$$\mathcal{R} = \{R_n(P) : n \in \mathbb{N}, P \in \Delta(S_n \times X), P_S = R_n(P)\} \tag{2.9}$$

Assumption 2 ensures that any sequence of stationary measures for the truncated transition probability functions $\{Q^n(s,x) : s \in S_n, x \in X\}$ is tight, which further implies that any sequence of stationary measures has a convergent subsequence. Tightness may sometimes be hard to verify directly and therefore, we provide two sufficient conditions that are satisfied for most applications.

**Condition 1 (Reversible)**: The transition probability function $\{Q(s,x)\}$ has a unique stationary measure $\pi$ and is *reversible* with respect to $\pi$. That is, there exists a unique $\pi \in \Delta(S \times X)$ such that

$$\pi_S(A) = \int_{S \times X} Q(s,x)(A)\pi(\mathrm{d}s, \mathrm{d}x) \tag{2.10}$$

for all $A \in \mathcal{B}[S]$. Moreover, for all $A_1, A_2 \in \mathcal{B}[S]$, we have

$$\int_{A_1 \times X} Q(s,x)(A_2)\pi(\mathrm{d}s, \mathrm{d}x) = \int_{A_2 \times X} Q(s,x)(A_1)\pi(\mathrm{d}s, \mathrm{d}x). \tag{2.11}$$

**Condition 2 (Lyapunov)**: The transition probability function $\{Q(s,x)\}$ satisfies the *Lyapunov condition*, that is, there exist a non-negative continuous norm-like function $V$,[14] and constants $0 < \alpha \leq 1$, $\beta \geq 0$ such that

$$\int_S V(y)Q(s,x)(\mathrm{d}y) \leq (1-\alpha)V(s) + \beta \tag{2.12}$$

[14]A function $V : S \to \mathbb{R}_{\geq 0}$ is *norm-like* if $\{s \in S : V(s) \leq B\}$ is precompact for every $B > 0$.

for all $s \in S$ and $x \in X$. Moreover, the sequence $(\{s \in S : V(s) \leq n\})_{n \in \mathbb{N}}$ of sets satisfies Assumption 1. Then, by taking $S_n = \{s \in S : V(s) \leq n\}$, Assumption 2 is satisfied.

To establish belief restriction for the full SMDP $\mathcal{M}$, we impose the following assumption on the relative entropy.

**Assumption 3.** (Uniform-integrability) For all $\theta \in \hat{\Theta}$, the family of relative entropy $\{\mathcal{D}_{\mathrm{KL}}\big(Q(s,x), Q_\theta(s,x)\big)\}$ is uniformly integrable with respect to all stationary $P \in \Delta(S \times X)$. That is, for every $\epsilon > 0$, there exists $\kappa > 0$ such that

$$\int_E \mathbb{E}_{Q(\cdot|s,x)} \left[\ln \left(D_\theta(s'|s,x)\right)\right] P(\mathrm{d}s, \mathrm{d}x) < \epsilon \tag{2.13}$$

for all $\theta \in \hat{\Theta}$ and all stationary $P \in \Delta(S \times X)$ with $P(E) < \kappa$.[15]

The candidate Berk-Nash equilibrium for $\mathcal{M}$ is the weak limit of Berk-Nash equilibrium for the sequence of truncated SMDPs. Assumption 3 allows us to approximate the weighted Kullback-Leibler divergence of $\mathcal{M}$ from the weighted Kullback-Leibler divergence of truncated SMDPs. Alternatively, we can also establish belief restriction under the following assumption.

**Assumption 4.** (Uniqueness) There exists a unique $\theta_0 \in \Theta$ that minimizes the relative entropy $\mathcal{D}_{\mathrm{KL}}\big(Q(s,x), Q_\theta(s,x)\big)$ for all $(s,x) \in S \times X$. Moreover, for every $n \in \mathbb{N}$, $\theta_0$ uniquely minimizes $\mathcal{D}_{\mathrm{KL}}\big(Q^n(s,x), Q^n_\theta(s,x)\big)$ for all $(s,x) \in S_n \times X$.

*Remark* 2.9. Under Assumption 4, the set of closest parameters for $\mathcal{M}$ and all truncated SMDPs is the same singleton set $\{\theta_0\}$. If the model is *correctly specified*, then Assumption 4 is trivially satisfied and the the set of closest parameters contains a single point, which is the true parameter value.

We establish optimality under two different set of conditions: Theorem 2 assumes the payoff function is bounded continuous while Theorem 3 allows for unbounded payoff functions under a fairly general norm-restriction assumption on the state space. For the bounded case, we assume,

**Assumption 5.** (Boundedness) The payoff function $\pi : S \times X \times S \to \mathbb{R}$ is a bounded continuous function.

[15]One sufficient condition for Assumption 3 is to assume that the relative entropy is uniformly bounded on the set $\{(\theta, s, x) \in \Theta \times S \times X : Q(s,x)$ is dominated by $Q_\theta(s,x)\}$. If the true transition probability function $Q$ and every element in $\mathcal{Q}_\Theta$ do not depend on the current state, as the action and parameter spaces are compact, this sufficient condition is usually satisfied.

We now present our second theorem that establishes the existence of Berk-Nash Equilibrium for bounded payoff functions with $\sigma$-locally compact state space.

**Theorem 2.** *Any regular SMDP that satisfies regularity (Assumption 1), tightness (Assumption 2) and has a bounded payoff function (Assumption 5) has a Berk-Nash equilibrium if either the SMDP is correctly specified or if one of the assumptions of uniform integrability (Assumption 3) or uniqueness (Assumption 4) holds.*

Unbounded payoff functions are common in many environments. We assume the state space $S$ is a norm space and use $\|s\|$ to denote the norm of an element $s \in S$. We impose the following assumption on the payoff function.

**Assumption 6.** (State-boundedness) The payoff function $\pi : S \times X \times S \to \mathbb{R}$ is a jointly continuous function and there exist $A, B \in \mathbb{R}_{>0}$ such that[16] for all $(s, x, s') \in S \times X \times S'$, $|\pi(s, x, s')| \leq A + B \max\{\|s\|, \|s'\|\}$.

Our final two assumptions are on the subjective transition probability functions.

**Assumption 7.** (Fold-boundedness) Let $B \in \mathbb{R}_{>0}$ be given in Assumption 6. There exist[17] some $C, D \in \mathbb{R}_{>0}$ such that $\int_S \|s'\| Q_\theta(\mathrm{d}s'|s, x) \leq C + \frac{D}{(1+\delta)B+\delta D}\|s\|$ for all $x \in X$, $\theta \in \Theta$ and $s \in S$.

The following stronger continuity condition is assumed on the family $\mathcal{Q}_\Theta = \{Q_\theta : \theta \in \Theta\}$. Let $d_S$ denote the metric on $S$ generated from the norm.

**Assumption 8.** (W-continuity) The mapping $(\theta, s, x) \to Q_\theta(s, x)$ is continuous in the 1-Wasserstein metric. Moreover, $Q_\theta(s, x)$ has finite first moment for all $(\theta, s, x) \in \Theta \times S \times X$. That is, for every $(\theta, s, x) \in \Theta \times S \times X$, there exists[18] some $s_0 \in S$ such that $\int_S d_S(t, s_0) Q_\theta(s, x)(\mathrm{d}t) < \infty$.

Our final theorem establishes existence for unbounded payoff functions.

**Theorem 3.** *Theorem 2 holds if the boundedness of the payoff function (Assumption 5) is weakened to state-boundedness of the payoff function (Assumption 6) but with subjective transition probability functions being fold-bounded (Assumption 7) and W-continuous (Assumption 8).*

[16] Instances where such an assumption is easily satisfied are common: (i) a monopolist's payoff in Nyarko (1991), (ii) CRRA payoff functions in stochastic growth and optimal savings environments.

[17] It is clear that $\frac{D}{(1+\delta)B+\delta D} \to \frac{1}{\delta}$ as $D \to \infty$. If we assume that there exist $C \in \mathbb{R}_{>0}$ and $D < \frac{1}{\delta}$ such that $\int_S \|s'\| Q_\theta(\mathrm{d}s'|s, x) \leq C + D\|s\|$, then it implies Assumption 7 with a suitably chosen $D'$.

[18] By the triangle inequality, under Assumption 8, we have $\int_S d_S(t, s') Q_\theta(s, x)(\mathrm{d}t) < \infty$ for all $s' \in S$ and all $(\theta, s, x) \in \Theta \times S \times X$.

## 3. Some Selected Examples

In this section, we present examples to demonstrate the applicability of our main results to a variety of problems encountered in economic theory. Examples 3.1 and 3.2 connect the existence of an equilibrium with the existence of unit roots for an AR(1) process. Examples 3.3, 3.4, and 3.5 illustrate settings that naturally involve continuous distributions.

**Example 3.1** (Unbounded Densities and Unit Root)**.** In this example, we show that for a AR(1) process, a Berk-Nash equilibrium exists if and only if the AR(1) process does not have a unit root. In addition, we illustrate that the unbounded density functions arise naturally in correctly specified econometric inference problems.

Consider a SMDP with state space $S = \mathbb{R}$, a singleton action space $X = \{0\}$, and a payoff function $\pi : S \times X \times S$ that equals 0 for all $(s, x, s) \in S \times X \times S$. For every $s \in S$, the true transition probability function $Q(s)$ is the distribution of $a_0 s + b_0 \xi$, where $a_0 \in [0, 2], b_0 \in [0, 1]$ and $\xi = \mathcal{N}(0, 1)$ has the standard normal distribution.[19] The parameter space $\Theta$ is $[0, 2] \times [0, 1]$ and for every $(a, b) \in \Theta$, the transition probability function $Q_{(a,b)}(s)$ is the distribution of $as + b\xi$.

Consider first the degenerate case $b_0 = 0$. In this case, the evolution of the state is deterministic. When $a_0 < 1$, the Dirac measure $\delta_{(0,0)}$ at $(0, 0)$ is a Berk-Nash equilibrium, supported by the belief $\delta_{(a_0,0)}$. When $a_0 = 1$, every Dirac measure $\delta_{(s,0)}$ for $s \in S$ is a Berk-Nash equilibrium supported by the belief $\delta_{(1,0)}$. There is no Berk-Nash equilibrium if $a_0 > 1$.

Now, we turn to the non-degenerate case $b_0 > 0$. The true transition probability function $Q(s) = Q_{(a_0,b_0)}(s)$ is absolutely continuous with respect to $Q_{(a,b)}(s)$ for all $(a, b) \in \hat{\Theta} = [0, 2] \times (0, 1]$, and the density function is jointly continuous function where it is defined. Note, however, that the density function is unbounded on $\hat{\Theta}$, tending to infinity as $b \to 0$. The unboundedness arises here, even though the model is correctly specified. This situation arises ubiquitously in econometric inference. In any OLS estimation, we test (among other things) whether or not the regression coefficient $\alpha_1$ of the dependent variable $y$ on a given independent variable $x_1$ is, or is not, zero. This estimation requires us to include in $\Theta$ the *possibility* that $\alpha_1$ is zero. If $\alpha_1$ is, in fact, not zero, the Radon-Nikodym derivative will typically be unbounded.

It is straightforward to verify that this is a regular SMDP in the sense of Definition 2.4. As the state space is not compact, we need to check the conditions under which this

[19] In other words, this is simply an inference problem about a Markov process, specifically an AR(1) process, rather than a full Markov *decision problem*.

example satisfies the assumptions for Theorem 2. We will see that the example satisfies those conditions if and only if $a_0 < 1$. Note that if $a_0 \geq 1$, then since $b_0 \neq 0$, the Markov process has no stationary distribution, and hence there is no Berk-Nash equilibrium. Note also that $a_0 \geq 1$ if and only if the AR(1) process has a unit root, which implies that the usual method for estimating the parameters of the AR(1), ordinary least squares, yields spurious results. Thus, our example has a Berk-Nash equilibrium if and only if the AR(1) process does not have a unit root.

To see this, suppose $0 \leq a_0 < 1$ (and recall that $b_0 > 0$). For each $n \in \mathbb{N}$, let $S_n = [-n, n]$; it is straightforward to see that the sequence $\{S_n : n \in \mathbb{N}\}$ satisfies Assumption 1. Assumption 2 is satisfied by taking the Lyapunov function $V(s) = |s|$. It can be verified that Assumption 3 is satisfied, which establishes belief restriction[20]. We can also establish belief restriction from the fact that the SMDP is correctly specified. We can easily modify this example to a SMDP with misspecification, in which case belief restriction follows from Assumption 4. The payoff function is constant, and hence satisfies Assumption 5. By Theorem 2, there exists a Berk-Nash equilibrium for this SMDP. The Berk-Nash equilibrium is $\mu \times \delta_0$ where $\mu = \mathcal{N}(0, \frac{b_0^2}{1-a_0^2})$, supported by the belief $\delta_{(a_0,b_0)}$.

**Example 3.2** (Markov Decision Problem with Unit Root)**.** We modify Example 3.1 by setting the action space $X = [0, 1]$ and $\Theta = [0, 2] \times [0, 1] \times [-1, 1]$. The true probability transition $Q(s, x)$ has the distribution of $a_0 s + b_0 \xi + c_0 x$, with $c_0 \in [-1, 1]$, and the payoff is $\pi(s, x, s') = s'$.[21] For every $(a, b, c) \in \Theta$, the transition probability function $Q_{(a,b,c)}(s, x) = as + b\xi + cx$. The degenerate case $b_0 = 0$ is handled in the same way as in Example 3.1.

Now suppose that $b_0 > 0$. If $a_0 \geq 1$, then since $b_0 \neq 0$, the Markov decision process has no stationary distribution, and hence there is no Berk-Nash equilibrium. If $a_0 < 1$, we restrict $\Theta$ to $\Theta' = [0, 1] \times [0, 1] \times [-1, 1]$, and verify that the Assumptions of Theorem 3 are satisfied for this modified SMDP. Letting $V(s) = \|s\|$ and $S_n = [-n, n]$, Eq. (2.12) is satisfied by essentially the same calculation as in Example 3.1. The payoff function $\pi$ clearly satisfies Assumption 6 and Assumption 8. By a similar calculation to Example 3.1, Assumption 3 is satisfied. We can establish belief restriction from the fact that this SMDP is correctly specified or via Assumption 4. By essentially the same calculation as in Example 3.1, Assumption 7 is satisfied. Thus, the restricted SMDP has a Berk-Nash equilibrium by Theorem 3.

[20]In Appendix C.4, we provide rigorous verification for Assumption 2 and Assumption 3.
[21]In contrast to Example 3.1, the action $x$ affects the evolution of the state.

When $b_0 > 0$ and $a_0 < 1$, observe that the action choice $x = \operatorname{sign} c_0$[22] is a dominant strategy,[23] so let $\mu$ be the unique stationary distribution on $S$ induced by the action choice $\operatorname{sign} c_0$. When $c_0 \neq 0$, the Berk-Nash equilibrium is $\mu \times \delta_{\operatorname{sign} c_0}$; when $c_0 = 0$, the set of Berk-Nash equilibria is $\mu \times \Delta(X)$; in both cases, the Berk-Nash equilibria are supported by the belief $\delta_{(a_0,b_0,c_0)}$. Note that, in this example, the set of closest parameter values $\Theta'_Q(\mu) = \{(a_0, b_0, c_0)\}$ for the restricted SMDP is the same as the set of closest parameter values $\Theta_Q(\mu)$ for the original SMDP. Hence, the equilibrium is a Berk-Nash equilibrium of the original SMDP. Therefore, by Theorem 3, the problem has a Berk-Nash equilibrium if and only if $a_0 < 1$, i.e. if and only if the problem does not have a unit root.

**Example 3.3** (Misspecified Revenue)**.** In this example, we incorporate misspecification in the payoff function with misspecified pricing shocks. The Markov Decision Process is as follows. Every period, an agent observes a productivity shock $z \in \mathbb{Z} = [0, 1]$ and chooses an input $x \in X \subset \mathbb{R}_+$ which results in the agent receiving a payoff of $r(x) - c(x)$, where $c(x) = x^2$ is the cost of choosing $x$, and $r(x) = zf(x)\epsilon$ where $f(x)$ is the production function, $\epsilon$ is a random, independent shock to the price (which we set as 1) distributed according to the (true) distribution $d^*$, which has support equal to $[0, b], 0 \leq b \leq \infty$ and $0 < \mathbb{E}_{d^*}[\epsilon] < \infty$.[24] Therefore, the state space is given by, $(z, \epsilon) \in S = [0, 1] \times [0, b]$. Let $Q\,(z' \mid z)$ be the probability that tomorrow's productivity shock is $z'$, given the current shock $z$ and similarly, let $Q^R(\epsilon' \mid x)$ denote the transition function for the price shock, $\epsilon'$. We follow EP in framing the price shock $\epsilon$ as a part of the state variables along with the productivity shock $z$ and define the Bellman equation below.

$$V(z, \epsilon) = \max_x \int_{[0,1]\times[0,b]} \left(zf(x)\epsilon' - c + \delta V\left(z', \epsilon'\right)\right) Q\left(\mathrm{d}z' \mid z\right) Q^R\left(\mathrm{d}\epsilon' \mid x\right) \tag{3.1}$$

We assume that there is a unique stationary distribution over these productivity shocks, denoted by $z \sim U[0, 1]$. Next, we describe the SMDP of our environment. The agent believes (SMDP) that $f(x) = x$ and $\epsilon \sim d_\theta$, where $d_\theta$ has support equal to $[0, b]$ where $b = k\theta$. The parameter space $\Theta$ and the action space $X$ are chosen as such to be compact.[25] We assume that $\epsilon$ follows a truncated exponential distribution, $d_\theta(\epsilon) = \dfrac{(1/\theta)e^{-(1/\theta)\epsilon}}{1 - e^{(-b/\theta)}}$. Here, the agent's model can be misspecified if either the true

[22] $\operatorname{sign} x = x/|x|$ when $x \neq 0$, $\operatorname{sign} 0 = 0$.
[23] It is weakly dominant if $c_0 = 0$, dominant otherwise.
[24] We assume that the true distribution $d^*$ satisfies conditions in Definition 2.4.
[25] The details are supplied in the Online Appendix.

production function is not linear or if the true distribution of revenue shocks are not a part of the exponential family, or if support assumed of the model transition functions is different from the true transtion function. Given these primitives, it is easy to verify that this is a regular SMDP in the sense of Definition 2.4. Therefore, from Theorem 1, a Berk-Nash equilibrium exists.

**Example 3.4** (Optimal Savings)**.** This example extends Example 2 of EP for a continuum of preference shocks in an optimal consumption-savings model. The Markov decision process consists of the following objects: A state space $S = (y, z) \in Y \times Z = (0, \infty) \times [0, 1]$, where $y$ and $z$ denote the wealth and preference shocks, respectively. For each $y \in Y$, the agent chooses $x \in X = [0, 1]$, with $x$ representing the *fraction* of $y$ the agent chooses to save, so that the agent saves $k = xy$ and consumes $y - k$.[26] The payoff function $\pi$ is $\pi\left(y, x, z\right) = z \ln\left(y - k\right) = z \ln\left(y - xy\right)$.[27] Given the distribution of the shocks, the agent maximizes their discounted expected utility by choosing optimal proportion of savings, $x$. We next describe the true transition function. $Q\left(y', z' \mid y, z, x\right)$ is such that $y'$ and $z'$ are independent, $y'$ has a log-normal distribution with mean $\alpha^* + \beta^* \ln(xy) + \gamma^* z$ and unit variance, and $z'$ is *uniform* on $[0, 1]$. That is, the next period wealth, $y_{t+1}$, is given by $\ln y_{t+1} = \alpha^* + \beta^* \ln x_t y_t + \varepsilon_t$, where $\varepsilon_t = \gamma^* z_t + \xi_t$ is an unobserved i.i.d. productivity shock, $\xi_t \sim N(0, 1), \gamma^* \neq 0$, and $0 \leq \beta^* < 1,\ \delta\beta^* < 1$, where $\delta \in [0, 1)$ is the discount factor.[28] The Bellman equation for this MDP is as follows.

$$V(y, z) = \max_{0 \leq x \leq 1} z \ln(y - xy) + \delta \mathbb{E}\left[V\left(y', z'\right) \mid x\right],$$

However, the agent believes (SMDP) that $\ln y_{t+1} = \alpha + \beta \ln(x_t y_t) + \varepsilon_t$ where $\varepsilon_t \sim N(0, 1)$ and is independent of the utility shock. Further, the agent knows the distribution of the utility shock and is uncertain about $\beta \in \Theta$ which is a compact set in $\mathbb{R}$. The *subjective* transition probability function $Q_\theta\left(y', z' \mid y, z, x\right)$ is such that $y'$ and $z'$ are independent, $y'$ has a log-normal distribution with mean $\alpha + \beta \ln(xy)$ and unit variance, and $z'$ is *uniform* on $[0, 1]$. The agent has a misspecified model because she believes that the productivity and utility shocks are independent, when in fact $\gamma^* \neq 0$. Here we

[26]In EP, the agent chooses how much $k \in [0, y]$ to save. Here, we recast the problem in terms of the fraction saved in order to ensure that the action set $X$ is compact and independent of the state.

[27]When $z = 0$, we again use the standard convention that $0 \ln 0 = 0$. When $z \neq 0$, we approximate the action space $X = [0, 1]$ by closed intervals $\{[0, 1 - \epsilon] : \epsilon > 0\}$.

[28]It is the restriction on $0 \leq \beta < 1$ that gives us stationarity. The detailed analysis is in the Supplementary Appendix.

diverge from the EP example by assuming that the preference shocks $z$ are distributed uniformly over $[0,1]$.

Following Examples 3.1 and 3.2, it is easy to see that the SMDP satisfies Definition 2.4 and given the normality of the transition probability function, Assumptions 1-3 hold. The state space is not compact, and therefore, we need to check whether Theorem 2 or Theorem 3 applies. The payoff function is unbounded. However, it is state-bounded and therefore, satisfies Assumption 6. Finally, Assumptions 7-8 hold as well as illustrated in the previous example. Therefore, by Theorem 3, a Berk-Nash equilibrium exists. We next characterize the Berk-Nash equilibrium for this instance.[29]

In this case, the Berk-Nash equilibrium is characterized by the optimal policy function, $k = x^* y = A_z(\beta^m) y = \frac{0.5\delta\beta^m}{(1-\delta\beta^m)z + 0.5\delta\beta^m} y$, where there exists a $\beta^m \in (0, \beta^*)$. Indeed, note that the true transition probability function $Q(s)$ has a unique stationary measure $\mu$. So, the Berk-Nash equilibrium for this SMDP is $\mu \times \delta_{x^*}$, supported by the belief $\delta_{(\beta^m)}$.

**Example 3.5** (Misspecified Costs)**.** Our final example extends EP's finite productivity shocks to a continuum of shocks in the realm of a producer's problem and closely mirrors Example 3.3, therefore the analysis is similar. However, instead of having unbounded support for the cost shock as in EP, we restrict it to come from a bounded support. Consider the following Markov decision process. Every period, an agent observes a productivity shock $z \in \mathbb{Z} = [0,1]$ and chooses an input $x \in X \subset \mathbb{R}_+$ which results in the agent obtaining a payoff of $r(x) - c(x)$ every period, where $c(x) = \phi(x)\epsilon$ is the cost of choosing $x$, $r(x) = z\ln(x)$ where $\ln(x)$ is the production function, $z$ is the productivity shock in $[0,1]$ and $\epsilon$ is a random, independent shock to the cost distributed according to the (true) distribution $d^*$, which has support equal to $[0,b], 0 \leq b \leq \infty$ and $0 < \mathbb{E}_{d^*}[\epsilon] < \infty$.[30] The state space $S = [0,1] \times [0,b], b < \infty$ is the support of the cost shock. The action space $X$ and the parameter space $\Theta$ is chosen as such to be compact[31] and the payoff function $\pi(s,x,s') = z\ln x - c(x)$. The Bellman equation is given by:

$$V(z,\epsilon) = \max_x \int_{[0,1]\times[0,b]} \left(zf(x) - c' + \delta V\left(z', \epsilon'\right)\right) Q\left(\mathrm{d}z' \mid z\right) Q^C\left(\mathrm{d}\epsilon' \mid x\right) \tag{3.2}$$

Let $Q\left(z' \mid z\right)$ be the probability that tomorrow's productivity shock is $z'$ given the current shock $z$. We assume that there is a unique stationary distribution over these

[29]The details are given in Appendix C.
[30]We assume that the true distribution $d^*$ satisfies conditions in Definition 2.4.
[31]The details are supplied in the Online Appendix.

productivity shocks which is uniform, $U[0,1]$. Similarly, let $Q^C(\epsilon' \mid x)$ denote the transition function for the cost shock, $\epsilon'$. The agent believes in a misspecified cost function (SMDP), that is, $c_\theta(x) = x\epsilon$ and $\epsilon \sim d_\theta$, where $d_\theta$ has support equal to $[0,b]$ where $b = k\theta$, $0 \le k < \infty$. We assume that $\epsilon$ follows a truncated exponential distribution, $d_\theta(\epsilon) = \dfrac{(1/\theta)e^{-(1/\theta)\epsilon}}{1-e^{(-b/\theta)}}$.[32] Given these primitives, it is easy to verify that this is a regular SMDP in the sense of Definition 2.4. Therefore, from Theorem 1, a Berk-Nash equilibrium exists.

## 4. The Nonstandard Framework

4.1. **Methodological Innovation.** While nonstandard analysis has been used in mathematical economics since the 1970s, this paper relies on a new nonstandard technique pioneered in Duanmu (2018) to extend theorems from finite mathematical structures to infinite mathematical structures.[33] Candidates for this technique have the following properties:

- The theorem is known on a finite (or finite-dimensional) space, *and*
- The theorem *statement* does not rely heavily on the space being finite, *but*
- The existing proof(s) *do* rely heavily on the space being finite.

For results with these properties, Duanmu's technique allows one to directly translate the *statement* of the theorem without having to translate the *details* of the proof.

*Nonstandard* models satisfy three principles: *extension*, which associates to every ordinary mathematical object a nonstandard counterpart called its extension; *transfer*, which preserves the truth values of first-order logic statements between standard and nonstandard models; and *saturation*, which gives us a powerful mechanism for proving the existence of nonstandard objects defined in terms of finitely satisfiable collections of first-order formulas. In a suitably saturated nonstandard model, one can construct a single object—a hyperfinite probability space—that satisfies all the first order logical properties of a finite probability space, but which can be simultaneously viewed as a measure-theoretical probability space via the Loeb measure construction.

Duanmu's technique invokes the following proof strategy:

[32] In the context of this particular example, the agent's model can be misspecified if either cost functions are nonlinear, true distribution of cost shocks are not a part of the exponential family, or if the support assumed is incorrect.

[33] This paper is part of an ongoing program applying nonstandard analysis to resolve important problems in Markov processes (Duanmu et al. (2021a), Anderson et al. (2018), Anderson et al. (2021b) and Anderson et al. (2021c)), statistics (Duanmu and Roy (2021) and Duanmu et al. (2021b)) and mathematical economics (Anderson et al. (2022) and Anderson et al. (2021a)).

- Start with an standard infinite (e.g. measure-theoretic) object.
- Construct a *lifting*, embedding our standard object in a hyperfinite object.
- Use the *transfer principle* to obtain the theorem for the hyperfinite object, essentially for free.
- Use the Loeb measure construction to *push down* the theorem for the hyperfinite object to obtain the result in the original standard setting.

The rest of the section proceeds as follows. After setting out some basic preliminaries of non-standard analysis for the lay reader in Section 4.2, we turn to an overview of the basic argumentation. Section 4.3 and Appendix C.2 start with a regular SMDP with compact state and action spaces, embed it in a hyperfinite SMDP, transfer existing results from EP to this hyperfinite SMDP, then conclude by a "push down" argument to obtain a Berk-Nash equilibrium. An analogous argument is presented in Section 4.4 for a regular SMDP with a $\sigma$-compact state space.

4.2. **Preliminaries on Nonstandard Analysis.** For those who are not familiar with nonstandard analysis, Anderson et al. (2021a) and Anderson et al. (2022) provide reviews tailored to economists. Cutland et al. (1995), Arkeryd et al. (1997), and Wolff and Loeb (2000) provide thorough introductions. We use $^*$ to denote the nonstandard extension map taking elements, sets, functions, relations, etc., to their nonstandard counterparts. In particular, ${}^*\mathbb{R}$ and ${}^*\mathbb{N}$ denote the nonstandard extensions of the reals and natural numbers, respectively. An element $r \in {}^*\mathbb{R}$ is *infinite* if $|r| > n$ for every $n \in \mathbb{N}$ and is *finite* otherwise. An element $r \in {}^*\mathbb{R}$ with $r > 0$ is *infinitesimal* if $r^{-1}$ is infinite. For $r, s \in {}^*\mathbb{R}$, we use the notation $r \approx s$ as shorthand for the statement "$|r - s|$ is infinitesimal," and similarly we use use $r \gtrapprox s$ as shorthand for the statement "either $r \geq s$ or $r \approx s$."

Given a topological space $(X, \mathcal{T})$, the monad of a point $x \in X$ is the set $\bigcap_{U \in \mathcal{T} : x \in U} {}^*U$. An element $x \in {}^*X$ is *near-standard* if it is in the monad of some $y \in X$. We say $y$ is the standard part of $x$ and write $y = \mathsf{st}(x)$. Note that such $y$ is unique provided that $X$ is a Hausdorff space. The *near-standard part* $\mathrm{NS}({}^*X)$ of ${}^*X$ is the collection of all near-standard elements of ${}^*X$. The standard part map $\mathsf{st}$ is a function from $\mathrm{NS}({}^*X)$ to $X$, taking near-standard elements to their standard parts. In both cases, the notation elides the underlying space $Y$ and the topology $\mathcal{T}$, because the space and topology will always be clear from context. For a metric space $(X, d)$, two elements $x, y \in {}^*X$ are *infinitely close* if ${}^*d(x, y) \approx 0$. An element $x \in {}^*X$ is near-standard if and only if it is infinitely close to some $y \in X$. An element $x \in {}^*X$ is finite if there exists $y \in X$ such that ${}^*d(x, y) < \infty$ and is infinite otherwise.

Let $X$ be a topological space endowed with Borel $\sigma$-algebra $\mathcal{B}[X]$ and let $\mathcal{M}(X)$ denote the collection of all finitely additive probability measures on $(X, \mathcal{B}[X])$. An internal probability measure $\mu$ on $({}^*X, {}^*\mathcal{B}[X])$ is an element of ${}^*\mathcal{M}(X)$. The Loeb space of the internal probability space $({}^*X, {}^*\mathcal{B}[X], \mu)$ is a countably additive probability space $({}^*X, \overline{{}^*\mathcal{B}[X]}, \overline{\mu})$ such that

$$\overline{{}^*\mathcal{B}[X]} = \{A \subset {}^*X | (\forall \epsilon > 0)(\exists A_i, A_o \in {}^*\mathcal{B}[X])(A_i \subset A \subset A_o \wedge \mu(A_o \setminus A_i) < \epsilon)\} \tag{4.1}$$

and

$$\overline{\mu}(A) = \sup\{\mathsf{st}(\mu(A_i)) | A_i \subset A, A_i \in {}^*\mathcal{B}[X]\} = \inf\{\mathsf{st}(\mu(A_o)) | A_o \supset A, A_o \in {}^*\mathcal{B}[X]\}. \tag{4.2}$$

Every standard model is closely connected to its nonstandard extension via the *transfer principle*, which asserts that a first order statement is true in the standard model if and only if it is true in the nonstandard model. Given a cardinal number $\kappa$, a nonstandard model is called $\kappa$-saturated if the following condition holds: let $\mathcal{F}$ be a family of internal sets, if $\mathcal{F}$ has cardinality less than $\kappa$ and $\mathcal{F}$ has the finite intersection property, then the total intersection of $\mathcal{F}$ is non-empty. In this paper, we assume our nonstandard model is as saturated as we need (see *e.g.* Arkeryd et al. (1997, Thm. 1.7.3) for the existence of $\kappa$-saturated nonstandard models for any uncountable cardinal $\kappa$).

The concept of "push-down," through which a standard object is constructed from a nonstandard object, is at the heart of nonstandard analysis and will be employed in the proofs of our theorems.

**Definition 4.1.** Let $Y$ be a Hausdorff space endowed with Borel $\sigma$-algebra $\mathcal{B}[Y]$. Let $P$ be an internal probability measure on $({}^*Y, {}^*\mathcal{B}[Y])$. The **push-down measure** of $P$ is defined to be a standard measure $P_p$ on $(Y, \mathcal{B}[Y])$ such that $P_p(A) = \overline{P}(\mathsf{st}^{-1}(A))$ for all $A \in \mathcal{B}[Y]$.

The following lemma from Duanmu and Roy (2021, Lemma. 6.1) shows that if the underlying space is compact, then the push-down of an internal probability measure is a standard probability measure.

**Lemma 4.2.** *Let $Y$ be a compact Hausdorff space endowed with Borel $\sigma$-algebra $\mathcal{B}[Y]$. Let $P$ be an internal probability measure on $({}^*Y, {}^*\mathcal{B}[Y])$. Then $P_p$ is a probability measure on $(Y, \mathcal{B}[Y])$.*

We first provide the existence result and then provide the existence result for the compact space. It is straightforward to verify that regular-SMDPs with finite state and action spaces, as defined in EP, are regular in the sense of Definition 2.4. EP's Theorem 1 proves the following result.

**Theorem 4.3** (EP). *Suppose $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ is a regular-SMDP such that*

*(1) The state space $S$ and the action space $X$ are both finite;*

*(2) The parameter space is a compact subset of Euclidean space.*

*Then there exists a Berk-Nash equilibrium.*

Note that every finite set can be embedded into a Euclidean space. Thus, the following *finite* result is an immediate consequence of Theorem 4.3.

**Lemma 4.4.** *Suppose $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ is a regular-SMDP such that the state space $S$, the action space $X$ and the parameter space $\Theta$ are finite. Then there exists a Berk-Nash equilibrium.*

4.3. **Existence of Equilibrium with Compact State Space.** In this section, we introduce a hyperfinite Markov decision process and use it to give a proof outline for Theorem 1. Throughout this section, We work with a regular SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ with a compact state space $S$. We first give the definition of a hyperfinite representation of compact metric spaces.

**Definition 4.5.** Let $(Y, d)$ be a compact metric space with Borel $\sigma$-algebra $\mathcal{B}[Y]$. A hyperfinite representation of $Y$ is a tuple $(T_Y, \{B_Y(t)\}_{t\in T_Y})$ such that

(1) $T_Y$ is a hyperfinite subset of ${}^*Y$ and $Y$ is a subset of $T_Y$;

(2) $t \in B_Y(t) \in {}^*\mathcal{B}[Y]$ for every $t \in T_Y$;

(3) For every $t \in T_Y$, the diameter of $B_Y(t)$ is infinitesimal;

(4) For every $t \in T_Y$, $B_Y(t)$ contains an ${}^*$open set;

(5) The hyperfinite collection $\{B_Y(t) : t \in T_Y)\}$ forms a ${}^*$partition of ${}^*Y$.

For every $y \in {}^*Y$, we use $t_y$ to denote the unique element in $T_Y$ such that $y \in B_Y(t_y)$.

The next result from Duanmu et al. (2021a, Thm. 6.6) guarantees the existence of a hyperfinite representation when the underlying space is a compact metric space.

**Lemma 4.6.** *Let $Y$ be a compact metric space with Borel $\sigma$-algebra $\mathcal{B}[Y]$. Then there exists a hyperfinite representation $(T_Y, \{B_Y(t)\}_{t\in T_Y})$ of $Y$.*

A hyperfinite Markov decision process is a ${}^*$Markov decision process where the state and action spaces are hyperfinite. We construct a hyperfinite Markov decision process (HMDP) from the Markov decision process $\langle S, X, q_0, Q, \pi, \delta\rangle$:

(1) Let $(T_S, \{B_S(s)\}_{s\in T_S})$ and $(T_X, \{B_X(x)\}_{x\in S_X})$ to be two hyperfinite representations of $S$ and $X$, respectively, as in Lemma 4.6. $T_S$ is the hyperfinite state space and $T_X$ is the hyperfinite action space;

(2) Define $h_0(\{s\}) = {}^*q_0(B_S(s))$ for every $s \in T_S$. Note that $h_0$ is an internal probability measure on $T_S$. $h_0$ denotes the initial distribution of states;

(3) For every $s, s' \in T_S, x \in T_X$, let $\mathbb{Q}(s,x)(s') = {}^*Q(s,x)(B_S(s'))$ and $\mathbb{Q}(s,x)(A) = \sum_{s'\in A}\mathbb{Q}(s,x)(s')$ for all internal $A \subset T_S$. We write $\mathbb{Q}(A|s,x)$ for $\mathbb{Q}(s,x)(A)$. Then, $\mathbb{Q} : T_S \times T_X \to {}^*\Delta(T_S)$ is an internal transition probability function;

(4) Define $\Pi : T_S \times T_X \times T_S \to {}^*\mathbb{R}$ to be the restriction of ${}^*\pi$ on $T_S \times T_X \times T_S$. $\Pi$ denotes the hyperfinite per-period payoff function;

(5) The discount factor $\delta$ remains the same as in Definition 2.1.

Every hyperfinite Markov Decision Process has the same first-order logic properties as a finite Markov Decision Process. By the transfer principle, ${}^*\mathscr{Q}_\Theta$ is an internal family of internal transition probability functions. Let $(T_\Theta, \{B_\Theta(\theta)\}_{\theta\in T_\Theta})$ be a hyperfinite representation of $\Theta$. By Definition 4.5, $B_\Theta(\theta)$ contains an ${}^*$open set for all $\theta \in T_\Theta$. Thus, we have $B_\Theta(\theta) \cap {}^*\hat{\Theta} \neq \emptyset$ for all $\theta \in T_\Theta$. So, without loss of generality, we can assume $T_\Theta \subset {}^*\hat{\Theta}$. For every $\theta \in T_\Theta$, every $s, s' \in T_S$ and every $x \in T_X$, define $\mathbb{Q}_\theta(s,x)(s') = {}^*Q_\theta(s,x)(B_S(s'))$ and let $\mathbb{Q}_\theta(s,x)(A) = \sum_{s'\in A}\mathbb{Q}_\theta(s,x)(s')$ for all internal $A \subset T_S$. We sometimes write $\mathbb{Q}_\theta(A|s,x)$ for $\mathbb{Q}_\theta(s,x)(A)$. The family $\mathscr{Q}_{T_\Theta} = \{\mathbb{Q}_\theta : \theta \in T_\Theta\}$ is an internal family of internal transition probability functions. The tuple $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ is a *Hyperfinite Subjective Markov Decision Process* (HSMDP), which has the same first-order logic properties as a finite SMDP. The agent's problem can be cast recursively as

$$\mathbb{V}(t) = \max_{x\in T_X} \sum_{s'\in T_S} \{\Pi(s,x,s') + \delta\mathbb{V}(s')\}\mathbb{Q}(s'|s,x) \tag{4.3}$$

where $\mathbb{V} : T_S \to \mathbb{R}$ is the unique solution to the hyperfinite Bellman equation.

**Definition 4.7.** An action $x$ is ${}^*$optimal given $s$ in the HMDP($\mathbb{Q}$) if

$$x \in \arg\max_{\hat{x}\in T_X} \sum_{s'\in T_S} \{\Pi(s,\hat{x},s') + \delta\mathbb{V}(s')\}\mathbb{Q}(s'|s,\hat{x}) \tag{4.4}$$

Furthermore, an action $x$ is S-optimal given $s$ in the HMDP($\mathbb{Q}$) if

$$\sum_{s' \in T_S} \{\Pi(s,x,s') + \delta \mathbb{V}(s')\}\mathbb{Q}(s'|s,x) \approx \max_{\hat{x} \in T_X} \sum_{s' \in T_S} \{\Pi(s,\hat{x},s') + \delta \mathbb{V}(s')\}\mathbb{Q}(s'|s,\hat{x}). \tag{4.5}$$

Clearly, ${}^*$optimality implies S-optimality. We now introduce the concept of hyperfinite weighted Kullback-Leibler divergence.

**Definition 4.8.** The hyperfinite weighted Kullback-Leibler divergence is a mapping $\mathbb{K}_{\mathbb{Q}} : {}^*\Delta(T_S \times T_X) \times T_\Theta \to {}^*\mathbb{R}_{\geq 0}$ such that for any $m \in {}^*\Delta(T_S \times T_X)$ and $\theta \in T_\Theta$:

$$\mathbb{K}_{\mathbb{Q}}(m,\theta) = \sum_{(s,x) \in T_S \times T_X} \mathbb{E}_{\mathbb{Q}(\cdot|s,x)} \left[ \ln \Big( \frac{\mathbb{Q}(s'|s,x)}{\mathbb{Q}_\theta(s'|s,x)} \Big) \right] m(\{(s,x)\}). \tag{4.6}$$

The set of closest parameter values given $m \in {}^*\Delta(T_S \times T_X)$ is the set

$$T_\Theta^{\mathbb{Q}}(m) = \arg\min_{\theta \in T_\Theta} \mathbb{K}_{\mathbb{Q}}(m,\theta). \tag{4.7}$$

The set of almost closest parameter values given $m \in {}^*\Delta(T_S \times T_X)$ is the external set

$$\hat{T}_\Theta^{\mathbb{Q}}(m) = \{\hat{\theta} \in T_\Theta : \mathbb{K}_{\mathbb{Q}}(m,\hat{\theta}) \approx \min_{\theta \in T_\Theta} \mathbb{K}_{\mathbb{Q}}(m,\theta)\}. \tag{4.8}$$

Note that $\mathbb{Q}_\theta(s'|s,x) = 0$ implies that $\mathbb{Q}(s'|s,x) = 0$ for all $\theta \in T_\Theta$, $(s,x,s') \in T_S \times T_X \times T_S$. Hence, the hyperfinite relative entropy, $\mathbb{E}_{\mathbb{Q}(\cdot|s,x)} \Big[ \ln \big( \frac{\mathbb{Q}(s'|s,x)}{\mathbb{Q}_\theta(s'|s,x)} \big) \Big]$, is well-defined. If $\mathbb{Q}(s'|s,x) = 0$, then the corresponding term is interpreted as 0. Note that the hyperfinite relative entropy is always non-negative. By transferring the finite existence result in Lemma 4.4, we have the following theorem.

**Theorem 4.9.** *The hyperfinite Markov decision process*

$$(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta \rangle, \mathscr{Q}_{T_\Theta}) \tag{4.9}$$

*has a hyperfinite Berk-Nash equilibrium. That is, there exists some* $m \in {}^*\Delta(T_S \times T_X)$ *and some hyperfinite belief* $\nu \in {}^*\Delta(T_\Theta)$ *such that*

*(1)* ***Optimality****: For all* $(s,x) \in T_S \times T_X$ *such that* $m(\{(s,x)\}) > 0$, *$x$ is ${}^*$optimal given $s$ in the HMDP($\bar{\mathbb{Q}}_\nu$), where* $\bar{\mathbb{Q}}_\nu = \sum_{\theta \in T_\Theta} \mathbb{Q}_\theta \nu(\{\theta\})$;
*(2)* ***Belief Restriction****: We have* $\nu \in {}^*\Delta(T_\Theta^{\mathbb{Q}}(m))$;
*(3)* ***Stationarity****:* $m_{T_S}(\{s'\}) = \sum_{(s,x) \in T_S \times T_X} \mathbb{Q}(s'|s,x) m(\{(s,x)\})$ *for all* $s' \in T_S$.

By Lemma 4.2, $m_p$ and $\nu_p$ are probability measures on $S \times X$ and $\Theta$, respectively.[34] To show that $m_p$ is a Berk-Nash equilibrium for the SMDP, we introduce the following

[34]These are the push-down of the internal probability measures.

definition of a Berk-Nash S-equilibrium, which is more general than hyperfinite Berk-Nash equilibrium.

**Definition 4.10.** An internal probability distribution $m \in {}^*\Delta(T_S \times T_X)$ is a Berk-Nash S-equilibrium of the hyperfinite SMDP$(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ if there exists a hyperfinite belief $\nu \in {}^*\Delta(T_\Theta)$ such that

(1) **S-Optimality**: For all $(s,x) \in T_S \times T_X$ such that $m(\{(s,x)\}) > 0$, $x$ is S-optimal given $s$ in the HMDP$(\bar{\mathbb{Q}}_\nu)$, where $\bar{\mathbb{Q}}_\nu = \sum_{\theta \in T_\Theta} \mathbb{Q}_\theta \nu(\{\theta\})$;
(2) **S-Belief Restriction**: The support of $\nu$ is a subset of $\hat{T}_\Theta^{\mathbb{Q}}(m)$;
(3) **S-Stationarity**: For all internal $A \subset T_S$:

$$m_{T_S}(A) \approx \sum_{(s,x) \in T_S \times T_X} \mathbb{Q}(A|s,x) m(\{(s,x)\}). \tag{4.10}$$

Theorem 4.11 then connects the Berk-Nash S-equilibrium to Berk-Nash equilibrium for the hyperfinite SMDP. The proof is provided in the Appendix.

**Theorem 4.11.** *Let $m \in {}^*\Delta(T_S \times T_X)$ be a Berk-Nash S-equilibrium for the hyperfinite SMDP $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ with the associated hyperfinite belief $\nu \in {}^*\Delta(T_\Theta)$. Then $m_p$ is a Berk-Nash equilibrium with the associated belief $\nu_p$.*

Finally, Theorem 1 is then an immediate consequence of Theorem 4.11.

***Proof of Theorem 1.*** By Theorem 4.9, let $m$ be the hyperfinite Berk-Nash equilibrium for the hyperfinite SMDP with the associated hyperfinite belief $\nu$. Clearly, every hyperfinite Berk-Nash equilibrium is a Berk-Nash S-equilibrium. The result then follows from Theorem 4.11 and this completes the proof. □

4.4. **Existence of Equilibrium with Sigma-Compact State Space.** In this section, we work with a regular SMDP $\mathcal{M} = (\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ but with a $\sigma$-compact state space and compact action and parameter spaces. As opposed to approximating $\mathcal{M}$ with truncations, nonstandard analysis provides an elegant alternative approach by using a single nonstandard SMDP with a "large" *compact state space to represent the $\mathcal{M}$. The nonstandard SMDP can also be viewed informally as the limiting object of a sequence of truncated SMDPs.

We extend the sequence $\{S_n\}_{n \in \mathbb{N}}$ in Assumption 1 to an internal sequence $\{{}^*S_n\}_{n \in {}^*\mathbb{N}}$. By the transfer principle, ${}^*S_n$ is a *compact set for all $n \in {}^*\mathbb{N}$. Pick some $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. As $\{S_n\}_{n \in \mathbb{N}}$ is a sequence of non-decreasing sets, we have ${}^*S_n \subset {}^*S_N$ for all $n \in \mathbb{N}$, which implies that $\mathrm{NS}({}^*S) \subset {}^*S_N$. In other words, the state space $S$ is a subset of ${}^*S_N$. The nonstandard subjective Markov decision process (NSMDP) $\mathcal{M}^N_{T_\Theta} = (\langle {}^*S_N, {}^*X, {}^*q_0^N, {}^*Q^N, {}^*\pi_N, \delta\rangle, {}^*\mathcal{Q}^N{}_{T_\Theta})$ is defined as:

(1) The *state space is ${}^*S_N$, endowed with *Borel $\sigma$-algebra ${}^*\mathcal{B}[{}^*S_N]$;
(2) The action space is ${}^*X$, endowed with *Borel $\sigma$-algebra ${}^*\mathcal{B}[{}^*X]$;
(3) The parameter space $T_\Theta$ is the hyperfinite representation of $\Theta$ chosen in Definition 4.5. Note that $T_\Theta \subset {}^*\hat{\Theta}$;
(4) ${}^*q_0^N(A) = \frac{{}^*q_0(A)}{{}^*q_0({}^*S_N)}$ for all $A \in {}^*\mathcal{B}[{}^*S_N]$;
(5) ${}^*Q^N : {}^*S_N \times {}^*X \to {}^*\Delta({}^*S_N)$ is the *transition probability function defined as ${}^*Q^N(s,x)(A) = \frac{{}^*Q(s,x)(A)}{{}^*Q(s,x)({}^*S_N)}$ for all $A \in {}^*\mathcal{B}[{}^*S_N]$;
(6) ${}^*\pi_n : {}^*S_N \times {}^*X \times {}^*S_N \to {}^*\mathbb{R}$ is the restriction of ${}^*\pi$ to ${}^*S_N \times {}^*X \times {}^*S_N$;
(7) the discounting factor $\delta$ remains the same;
(8) For every $\theta \in T_\Theta$, ${}^*Q_\theta^N : {}^*S_N \times {}^*X \to {}^*\Delta({}^*S_N)$ is the *transition probability function defined as ${}^*Q_\theta^N(s,x)(A) = \frac{{}^*Q_\theta(s,x)(A)}{{}^*Q_\theta(s,x)({}^*S_N)}$ for all $A \in {}^*\mathcal{B}[{}^*S_N]$. Let ${}^*\mathcal{Q}^N{}_{T_\Theta} = \{{}^*Q_\theta^N : \theta \in T_\Theta\}$.

Under Assumption 1, every truncation of $\mathcal{M}$ is a regular SMDP and has a Berk-Nash equilibrium. Therefore, from the transfer principle, we have the following result.

**Theorem 4.12.** *Suppose Assumption 1 holds. Then $\mathcal{M}^N_{T_\Theta}$ is *regular and has a hyperfinite Berk-Nash equilibrium.*

***The Proofs of Theorem 2 and Theorem 3.*** By Theorem 4.12, $\mathcal{M}^N_{T_\Theta}$ has a Berk-Nash *equilibrium $m$ with the associated *belief function $\nu$ on ${}^*\Theta$. By Assumption 2, $m_p$ is a probability measure on $S \times X$. Let $m$ be the hyperfinite Berk-Nash equilibrium for $\mathcal{M}^N_{T_\Theta}$ with the associated hyperfinite belief $\nu \in {}^*\Delta(T_\Theta)$. Then, we have

$$m_{{}^*S}(A) = m_{{}^*S_N}(A) = \int_{{}^*S_N \times {}^*X} {}^*Q^N(A|s,x) m(\mathrm{d}s, \mathrm{d}x) \tag{4.11}$$

for all $A \in {}^*\mathcal{B}[{}^*S_N]$. Thus, $m$ is an element of ${}^*\mathcal{R}$, where $\mathcal{R}$ is the set in Assumption 2. Under Assumption 2, $\overline{m_{{}^*S}}(\mathsf{st}^{-1}(S)) = 1$, hence the push down $m_p$ is a probability measure on $S \times X$. As $\Theta$ is compact, by Lemma 4.2, $\nu_p$ is a probability measure on $\Theta$. In order to prove Theorems 2 and 3, therefore, it is sufficient to show that $m_p$ is a Berk-Nash equilibrium for the regular SMDP $\mathcal{M}$ with the belief $\nu_p$ on $\Theta$. The stationarity of $m_p$ follows from Theorem A.27 in the Appendix. In a series of Theorems A.31-A.33, we establish belief restriction for $\nu_p$ under the assumptions of a correctly specified SMDP, uniform integrability (Assumption 3) and uniqueness (Assumption 4), respectively. Finally, for a bounded payoff function, optimality follows from Theorem A.37, thus proving Theorem 2. For unbounded payoff function as considered in Assumption 6, optimality follows from Theorem A.45, hence proving Theorem 3. □

## 5. Discussion

This paper uses a novel technique in nonstandard analysis to extend the existence results for Berk-Nash equilibrium from finite state and action spaces to sigma-compact state and compact action spaces, thereby allowing coverage of a wide range of natural examples in macroeconomics, microeconomics, and finance. This paper suggests the following promising directions for future work:

(1) The paper, like EP, considers a single-agent environment. In future work, we hope to extend these results to the case of multiple agents, in particular recursive equilibrium frameworks in macroeconomics Molavi (2019);

(2) Theorem 4 in the online appendix provides a possible learning foundation for SMDPs with compact state and action spaces. Unfortunately, it relies on an implausibly strong condition, convergence in the total variation norm on measures. It is of great interest to develop a learning foundation under a weaker convergence condition such as convergence in the Prokhorov metric. This may have further implications for environments that are characterized by *slow* learning as in Frick et al. (2020).

## A. Appendix A. Proofs

In this appendix, we present proofs that are omitted from the main body of the paper. Most of the proofs make heavy use of nonstandard analysis for which notations are introduced in Section 4.2.

### A.1. **Proof of Theorem 1.**

In this appendix, we provide a rigorous proof to Theorem 4.11, thereby completing the proof of Theorem 1. The following two lemmas are key to prove the existence of a hyperfinite Berk-Nash equilibrium in Theorem 4.9. The first lemma follows from the fact that $T_\Theta$ is hyperfinite.

**Lemma A.1.** *Suppose* $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ *is a regular-SMDP. Then, for all* $(s, x, s') \in T_S \times T_X \times T_S$*, the function* $\mathbb{Q}_\theta(s'|s, x)$ *is* ${}^*$*continuous function of* $\theta$.

**Lemma A.2.** *Suppose* $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ *is a regular SMDP. Then, for all* $\theta \in T_\Theta$, $\mathbb{Q}_\theta(s'|s, x) > 0$ *for all* $(s, s', x) \in T_S \times T_S \times T_X$ *such that* $\mathbb{Q}(s'|s, x) > 0$.

*Proof.* Pick some $\theta \in T_\Theta$ and $(s, s', x) \in T_S \times T_S \times T_X$ with $\mathbb{Q}(s'|s,x) > 0$. Note that $\mathbb{Q}(s'|s,x) = {}^*Q(s,x)(B_S(s'))$. As $T_\Theta \subset {}^*\hat{\Theta}$, by the transfer principle, we have ${}^*Q_\theta(s,x)(B_S(s')) > 0$. As $\mathbb{Q}_\theta(s'|s,x) = {}^*Q_\theta(s,x)(B_S(s'))$, we have the result. □

***Proof of Theorem 4.9.*** Note that $T_\Theta$ is a hyperfinite set. Then the result follows from Lemma A.1, Lemma A.2 and the transfer of Lemma 4.4. □

Next, we divide the proof of Theorem 4.11 into the three subsections (A.1.1-A.1.3), thus establishing stationarity, optimality and belief restriction of the candidate Berk-Nash equilibrium $(m_p, \nu_p)$, respectively.

A.1.1. ***Stationarity.*** Recall that $(m_p)_S$ denotes the marginal measure of $m_p$ on $S$. In this section, we establish the stationarity of $(m_p)_S$. We use $m_{T_S}$ to denote the marginal measure of $m$ on $T_S$.

**Lemma A.3.** *For any $A \in \mathcal{B}[S]$, $(m_{T_S})_p(A) = (m_p)_S(A)$.*

*Proof.* We have $(m_{T_S})_p(A) = \overline{m_{T_S}}(\mathsf{st}^{-1}(A) \cap T_S) = \overline{m}\big((\mathsf{st}^{-1}(A) \cap T_S) \times T_X\big)$ for every $A \in \mathcal{B}[S]$. On the other hand, we have $(m_p)_S(A) = m_p(A \times X) = \overline{m}\big((\mathsf{st}^{-1}(A) \cap T_S) \times T_X\big)$ for all $A \in \mathcal{B}[S]$. Hence, we have the desired result. □

**Lemma A.4.** *Let $A$ be a (possibly external) subset of $T_S$. Suppose there exists a sequence $\{A_k : k \in \mathbb{N}\}$ of non-decreasing internal subsets of $T_S$ such that $\bigcup_{k\in\mathbb{N}} A_k = A$. Then $\overline{m_{T_S}}(A) = \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(A \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x)$*

*Proof.* By the continuity of probability, we have $\overline{m_{T_S}}(A) = \lim_{k\to\infty} \overline{m_{T_S}}(A_k)$. For each $k \in \mathbb{N}$, by the S-stationarity of $m$, we have

$$\overline{m_{T_S}}(A_k) \approx \int_{T_S\times T_X} \mathbb{Q}(s,x)(A_k)m(\mathrm{d}s, \mathrm{d}x) \approx \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(A_k)\overline{m}(\mathrm{d}s, \mathrm{d}x). \tag{A.1}$$

Thus, we have $\overline{m_{T_S}}(A) = \lim_{k\to\infty} \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(A_k)\overline{m}(\mathrm{d}s, \mathrm{d}x)$. The result then follows from the dominated convergence theorem. □

To complete the proof, we need to make a topological assumption on $S$. We start with the following definition.

**Definition A.5.** A $\pi$-system on a set $\Omega$ is a non-empty collection $P$ of subsets of $\Omega$ that is closed under finite intersection.

**Lemma A.6** (The Uniqueness Lemma)**.** *Let $(\Omega, \Sigma)$ be a measure space with $\Sigma$ generated from some $\pi$-system $\Pi$. Let $\mu$ and $\nu$ be two probability measures that agree on $\Pi$. Then $\mu$ and $\nu$ agree on $\Sigma$.*

**Assumption 9.** There exists a $\pi$-system $\mathcal{F}$ on $S$ that generates $\mathcal{B}[S]$ such that, for every $A \in \mathcal{F}$, $\mathsf{st}^{-1}(A) = \bigcup_{k\in\mathbb{N}} A_k$ for some sequence $\{A_k : k \in \mathbb{N}\} \subset {}^*\mathcal{B}[{}^*S]$ of (non-decreasing) sets.

Although Assumption 9 is stated in nonstandard terminology, it is satisfied by many standard topological spaces. In fact, all metric spaces which are endowed with the Borel $\sigma$-algebra satisfy Assumption 9.

**Theorem A.7.** *Let $Y$ be a metric space endowed with the Borel $\sigma$-algebra $\mathcal{B}[Y]$. Then $(Y, \mathcal{B}[Y])$ satisfies Assumption 9.*

*Proof.* Let $\mathcal{F}$ be the $\pi$-system generated by the collection of open balls. Clearly, $\mathcal{F}$ generates $\mathcal{B}[Y]$. Let $B(a, \eta)$ be an open ball centered at $a$ with radius $\eta$. For each $n \in \mathbb{N}$, let $C_n$ be the closure of $B(a, \eta - \frac{1}{n})$. Then, we have $\mathsf{st}^{-1}\big(B(a,\eta)\big) = \bigcup_{n\in\mathbb{N}} {}^*C_n$. Pick some $U \in \mathcal{F}$. Then $U = \bigcap_{i\leq n} U_i$ for some $n \in \mathbb{N}$, where $U_i$ is an open ball for all $i \leq n$. For each $i \leq n$, there is a sequence $\{A^i_k : k \in \mathbb{N}\} \subset {}^*\mathcal{B}[{}^*Y]$ such that $\mathsf{st}^{-1}(U_i) = \bigcup_{k\in\mathbb{N}} A^i_k$. Then $U$ equals to the union of the countable collection $\{\bigcap_{i\leq n} A^i_{k_i} : k_1, k_2, \ldots, k_n \in \mathbb{N}\}$. ☐

**Lemma A.8.** $(m_p)_S(A) = \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(\mathsf{st}^{-1}(A) \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x)$ *for all $A \in \mathcal{B}[S]$.*

*Proof.* By Theorem A.7, let $\mathcal{F}$ denote the $\pi$-system in Assumption 9. By Lemma A.3, we have $(m_p)_S(A) = (m_{T_S})_p(A) = \overline{m_{T_S}}(\mathsf{st}^{-1}(A) \cap T_S)$ for every $A \in \mathcal{B}[S]$. Pick some $B \in \mathcal{F}$. By Assumption 9, there is a sequence $\{B_k : k \in \mathbb{N}\} \subset {}^*\mathcal{B}[{}^*S]$ of non-decreasing sets such that $\mathsf{st}^{-1}(B) = \bigcup_{k\in\mathbb{N}} B_k$. By Lemma A.4, we have

$$(m_p)_S(B) = \overline{m_{T_S}}(\mathsf{st}^{-1}(B) \cap T_S) = \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(\mathsf{st}^{-1}(B) \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x). \qquad \text{(A.2)}$$

For every $A \in \mathcal{B}[S]$, define $P(A) = \int_{T_S\times T_X} \overline{\mathbb{Q}(s,x)}(\mathsf{st}^{-1}(A) \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x)$. It is straightforward to verify that $P$ is a well-defined a probability measure on $(S, \mathcal{B}[S])$. As $(m_p)_S$ and $P$ agree on $\mathcal{F}$, by Lemma A.6, we have the desired result. ☐

Next, we quote the following results from nonstandard analysis which will be used for the subsequent proofs.

**Theorem A.9** (Anderson (1982, Prop. 8.4))**.** *Let $Y$ be a compact Hausdorff space endowed with Borel $\sigma$-algebra $\mathcal{B}[Y]$, let $\nu$ be an internal probability measure on $({}^*Y, {}^*\mathcal{B}[Y])$, and let $f : Y \to \mathbb{R}$ be a bounded measurable function. Define $g : {}^*Y \to \mathbb{R}$ by $g(s) = f(\mathsf{st}(s))$. Then we have $\int f(y)\nu_p(\mathrm{d}y) = \int g(y)\overline{\nu}(\mathrm{d}y)$.*

**Theorem A.10** (Anderson and Rashid (1978, Corollary. 5))**.** *Let $Y$ be a compact Hausdorff space endowed with Borel $\sigma$-algebra $\mathcal{B}[Y]$, let $\{P_n\}_{n\in\mathbb{N}}$ be a sequence of probability measures on $(Y, \mathcal{B}[Y])$. Then the sequence $\{P_n\}_{n\in\mathbb{N}}$ converges weakly to a probability measure $P$ on $(Y, \mathcal{B}[Y])$ if and only if $P(A) = \overline{{}^*P_N}(\mathsf{st}^{-1}(A))$ for all $A \in \mathcal{B}[Y]$ and $N \in {}^*\mathbb{N} \setminus \mathbb{N}$.*

Recall that we assume the mappings $(s, x) \to Q(s, x)$ and $(\theta, s, x) \to Q_\theta(s, x)$ are continuous in the Prokhorov metric. By Theorem A.10, we have the following result:

**Lemma A.11.** *For every $(s, x) \in T_S \times T_X$, every $\theta \in T_\Theta$ and every $A \in \mathcal{B}[S]$, we have $Q(\mathsf{st}(s), \mathsf{st}(x))(A) = \overline{\mathbb{Q}(s, x)}(\mathsf{st}^{-1}(A) \cap T_S)$ and $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))(A) = \overline{\mathbb{Q}_\theta(s, x)}(\mathsf{st}^{-1}(A) \cap T_S)$.*

*Proof.* Pick $(s_0, x_0) \in T_S \times T_X$, $\theta_0 \in T_\Theta$ and $A_0 \in \mathcal{B}[S]$. By Theorem A.10, we have $Q(\mathsf{st}(s_0), \mathsf{st}(x_0))(A_0) = \overline{{}^*Q(s_0, x_0)}(\mathsf{st}^{-1}(A_0))$ and $Q_{\mathsf{st}(\theta_0)}(\mathsf{st}(s_0), \mathsf{st}(x_0))(A_0) = \overline{{}^*Q_{\theta_0}(s_0, x_0)}(\mathsf{st}^{-1}(A_0))$. As $\mathsf{st}^{-1}(A_0) = \bigcup\{B_S(s) : s \in \mathsf{st}^{-1}(A_0) \cap T_S\}$, by construction, we obtain the desired result. ◻

We now prove the main result of this section:

**Theorem A.12.** $(m_p)_S(A) = \int_{S\times X} Q(A|s, x) m_p(\mathrm{d}s, \mathrm{d}x)$ *for every $A \in \mathcal{B}[S]$.*

*Proof.* By Lemma A.8, we have $(m_p)_S(A) = \int_{T_S\times T_X} \overline{\mathbb{Q}(s, x)}(\mathsf{st}^{-1}(A) \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x)$ for all $A \in \mathcal{B}[S]$. Thus, it is sufficient to show that

$$\int_{T_S\times T_X} \overline{\mathbb{Q}(s, x)}(\mathsf{st}^{-1}(A) \cap T_S)\overline{m}(\mathrm{d}s, \mathrm{d}x) = \int_{S\times X} Q(s, x)(A) m_p(\mathrm{d}s, \mathrm{d}x). \quad \text{(A.3)}$$

This follows from Theorem A.9 and Lemma A.11. ◻

A.1.2. ***Belief Restriction.*** Recall that $\nu$ is the hyperfinite belief as in Theorem 4.9. As $\Theta$ is compact, $\nu_p$ is a well-defined probability measure on $\Theta$. In this section, we show that the support of $\nu_p$ is a subset of $\Theta_Q(m_p)$. We start with the following result, which is closely related to Zimmer (2005), on hyperfinite representation of density functions.

**Theorem A.13.** *For all $\theta \in T_\Theta$, all $(s, x, s') \in T_S \times T_X \times T_S$ such that*

*(1)* $\mathbb{Q}_\theta(s'|s, x) > 0$;
*(2)* $Q(\mathsf{st}(s), \mathsf{st}(x))$ *is dominated by* $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))$;
*(3)* $D_{\mathsf{st}(\theta)}(\mathsf{st}(s')|\mathsf{st}(s), \mathsf{st}(x))$ *is finite.*

*Then, we have* $\frac{\mathbb{Q}(s'|s,x)}{\mathbb{Q}_\theta(s'|s,x)} \approx D_{\mathsf{st}(\theta)}(\mathsf{st}(s')|\mathsf{st}(s), \mathsf{st}(x))$.

*Proof.* Pick some $\theta_0 \in T_\Theta$, some $(s_0, x_0, s_0') \in T_S \times T_X \times T_S$ that satisfy the assumptions of the theorem. As $T_\Theta \subset {}^*\hat{\Theta}$, by the transfer principle, we have

$$\mathbb{Q}(s_0'|s_0, x_0) = {}^*Q(s_0, x_0)(B_S(s_0')) = \int_{B_S(s_0')} {}^*D_{\theta_0}(y|s_0, x_0)^*Q_{\theta_0}(\mathrm{d}y|s_0, x_0). \tag{A.4}$$

We also have $\mathbb{Q}(s_0'|s_0, x_0) = \int_{B_S(s_0')} \frac{\mathbb{Q}(s_0'|s_0,x_0)}{\mathbb{Q}_{\theta_0}(s_0'|s_0,x_0)} {}^*Q_{\theta_0}(\mathrm{d}y|s_0, x_0)$. Note that the density function $D_\theta(s'|s, x)$ is jointly continuous on $\{(\theta, s', s, x) : Q(s, x) \text{ is dominated by } Q_\theta(s, x)\}$ and $D_{\mathsf{st}(\theta_0)}(\mathsf{st}(s_0')|\mathsf{st}(s_0), \mathsf{st}(x_0))$ is finite. Thus, we have ${}^*D_{\theta_0}(y|s_0, x_0) \approx \frac{\mathbb{Q}(s_0'|s_0,x_0)}{\mathbb{Q}_{\theta_0}(s_0'|s_0,x_0)}$ for all $y \in B_S(s_0')$. Hence, we conclude that $\frac{\mathbb{Q}(s_0'|s_0,x_0)}{\mathbb{Q}_{\theta_0}(s_0'|s_0,x_0)} \approx D_{\mathsf{st}(\theta_0)}(\mathsf{st}(s_0')|\mathsf{st}(s_0), \mathsf{st}(x_0))$, completing the proof. □

We now introduce the notion of S-integrability from nonstandard analysis.

**Definition A.14.** Let $(\Omega, \mathcal{A}, P)$ be an internal probability space and let $F : \Omega \to {}^*\mathbb{R}$ be an internally integrable function such that $\mathsf{st}(F)$ exists $\overline{P}$-almost surely. Then $F$ is *S-integrable* with respect to $P$ if $\mathsf{st}(F)$ is $\overline{P}$-integrable, and $\int |F|\mathrm{d}P \approx \int \mathsf{st}(|F|)\mathrm{d}\overline{P}$.

We now show that the hyperfinite Kullback-Leibler divergence is infinitely close to the standard Kullback-Leibler divergence. Recall that $\Theta_m = \{\theta \in \Theta : K_Q(m, \theta) < \infty\}$ for $m \in \Delta(S \times X)$. Note that $\hat{\Theta} \subset \Theta_m$ for all $m \in \Delta(S \times X)$.

**Theorem A.15.** *Let $\lambda$ be an element of ${}^*\Delta(T_S \times T_X)$. Then, we have*

*(1) $\mathbb{K}_{\mathbb{Q}}(\lambda, \theta) \gtrapprox K_Q(\lambda_p, \mathsf{st}(\theta))$ for all $\theta \in T_\Theta$ such that $\mathsf{st}(\theta) \in \Theta_{\lambda_p}$;*

*(2) $\mathbb{K}_{\mathbb{Q}}(\lambda, \theta) \approx K_Q(\lambda_p, \mathsf{st}(\theta))$ for all $\theta \in T_\Theta$ such that $\mathsf{st}(\theta) \in \hat{\Theta}$.*

*Proof.* Pick $\theta \in T_\Theta$ such that $\mathsf{st}(\theta) \in \Theta_{\lambda_p}$. As $K_Q(\lambda_p, \mathsf{st}(\theta)) < \infty$, this implies that $Q(s, x)$ is dominated by $Q_{\mathsf{st}(\theta)}(s, x)$ for $\lambda_p$-almost all $(s, x) \in S \times X$. The proof of the theorem relies essentially on the following claim which is proved in the supplementary material, C.1.

**Claim A.16.** *For every $(s, x) \in T_S \times T_X$ such that $Q(\mathsf{st}(s), \mathsf{st}(x))$ is dominated by $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))$, $\mathbb{E}_{\mathbb{Q}(\cdot|s,x)}\left[\ln\left(\frac{\mathbb{Q}(s'|s,x)}{\mathbb{Q}_\theta(s'|s,x)}\right)\right] \approx \mathbb{E}_{Q(\cdot|\mathsf{st}(s),\mathsf{st}(x))}\left[\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s), \mathsf{st}(x))\right)\right]$.*

Define $g : S \times X \to \mathbb{R}$ to be $g(s, x) = \mathbb{E}_{Q(\cdot|s,x)}\left[\ln\left(D_{\mathsf{st}(\theta)}(s'|s, x)\right)\right]$ for all $(s, x) \in S \times X$ such that $Q(s, x)$ is dominated by $Q_{\mathsf{st}(\theta)}(s, x)$ and $g(s, x) = 0$ otherwise. For each $n \in \mathbb{N}$, define $g_n : S \times X \to \mathbb{R}$ to be $g_n(s, x) = \min\{g(s, x), n\}$. As $K_Q(\lambda_p, \mathsf{st}(\theta)) < \infty$, we conclude that $K_Q(\lambda_p, \mathsf{st}(\theta)) = \lim_{n\to\infty} \int g_n(s, x)\lambda_p(\mathrm{d}s, \mathrm{d}x)$. Note that each $g_n$ is a bounded measurable function. Similarly, we define $G : T_S \times T_X \to {}^*\mathbb{R}$ to be $G(s, x) = \mathbb{E}_{\mathbb{Q}(\cdot|s,x)}\left[\ln\left(\frac{\mathbb{Q}(s'|s,x)}{\mathbb{Q}_\theta(s'|s,x)}\right)\right]$. For each $n \in \mathbb{N}$, let $G_n : T_S \times$

$T_X \to {}^*\mathbb{R}$ be $G_n(s,x) = \min\{G(s,x), n\}$. By Claim A.16 and Theorem A.9, we have $\int_{T_S\times T_X} G_n(s,x)\lambda(\mathrm{d}s,\mathrm{d}x) \approx \int_{S\times X} g_n(s,x)\lambda_p(\mathrm{d}s,\mathrm{d}x)$ for all $n\in\mathbb{N}$. Note that $\mathbb{K}_{\mathbb{Q}}(\lambda,\theta) \geq \lim_{n\to\infty}\int_{T_S\times T_X} G_n(s,x)\lambda(\mathrm{d}s,\mathrm{d}x)$. Thus, we have $\mathbb{K}_{\mathbb{Q}}(\lambda,\theta) \gtrapprox K_Q(\lambda_p, \mathsf{st}(\theta))$.

For the special case that $\mathsf{st}(\theta)\in\hat{\Theta}$, by Arkeryd et al. (1997, Section 4, Corollary 6.1) and Claim A.16, $G$ is S-integrable with respect to $\lambda$. Then, $\mathbb{K}_{\mathbb{Q}}(\lambda,\theta) \approx \lim_{n\to\infty}\int_{T_S\times T_X} G_n(s,x)\lambda(\mathrm{d}s,\mathrm{d}x)$ follows from Arkeryd et al. (1997, Section 4, Theorem 6.2). Hence, $\mathbb{K}_{\mathbb{Q}}(\lambda,\theta)\approx K_Q(\lambda_p,\mathsf{st}(\theta))$ when $\mathsf{st}(\theta)\in\hat{\Theta}$. □

We now prove the main result of this section.

**Theorem A.17.** *The support of $\nu_p$ is a subset of $\Theta_Q(m_p)$.*

*Proof.* Pick $\theta_0\in\Theta$ such that $\theta_0$ is in the support of $\nu_p$. As $\nu_p(A) = \overline{\nu}(\mathsf{st}^{-1}(A))$ for all $A\in\mathcal{B}[\Theta]$, by Theorem 4.9, there exists $\theta_1\approx\theta_0$ such that $\theta_1\in\hat{T}^{\mathbb{Q}}_{\Theta}(m)$. That is, we have $\mathbb{K}_{\mathbb{Q}}(m,\theta_1)\approx\min_{\theta\in T_\Theta}\mathbb{K}_{\mathbb{Q}}(m,\theta)$. Suppose there exists $\theta'\in\Theta$ such that $K_Q(m_p,\theta') < K_Q(m_p,\theta_0) - \frac{1}{n}$ for some $n\in\mathbb{N}$. Note that $K_Q(m_p,\theta)$ is a continuous function of $\theta$ on $\Theta_{m_p}$. As $\hat{\Theta}\subset\Theta_{m_p}$ and $\hat{\Theta}$ is a dense subset of $\Theta$, there exists some $\hat{\theta}\in\hat{\Theta}$ such that $K_Q(m_p,\hat{\theta}) < K_Q(m_p,\theta_0) - \frac{1}{2n}$. Let $t_{\hat{\theta}}\in T_\Theta$ be the unique element such that $\hat{\theta}\in B_\Theta(t_{\hat{\theta}})$. By Theorem A.15, we have

$$\mathbb{K}_{\mathbb{Q}}(m,t_{\hat{\theta}}) \approx K_Q(m_p,\hat{\theta}) < K_Q(m_p,\theta_0) - \frac{1}{2n} \lessapprox \mathbb{K}_{\mathbb{Q}}(m,\theta_1) - \frac{1}{2n}. \tag{A.5}$$

This is a contradiction, so the support of $\nu_p$ is a subset of $\Theta_Q(m_p)$. □

A.1.3. ***Optimality***. In this section, we establish the optimality of the candidate Berk-Nash equilibrium $m_p$.

**Lemma A.18.** *For every $\lambda\in{}^*\Delta(T_\Theta)$ and every $(t,x)\in T_S\times T_X$, $(\mathbb{Q}_\lambda(t,x))_p = Q_{\lambda_p}(\mathsf{st}(t),\mathsf{st}(x))$. That is, the push-down of $\mathbb{Q}_\lambda(t,x)$ is the same as $Q_{\lambda_p}(\mathsf{st}(t),\mathsf{st}(x))$.*

*Proof.* Fix $\lambda\in{}^*\Delta(T_\Theta)$ and $(t,x)\in T_S\times T_X$. Pick $A\in\mathcal{B}[S]$. By the construction of the Loeb measure, we have

$$(\mathbb{Q}_\lambda(t,x))_p(A) = \overline{\mathbb{Q}_\lambda(t,x)}(\mathsf{st}^{-1}(A)\cap T_S) = \int_{S_\Theta}\overline{\mathbb{Q}_i(t,x)}(\mathsf{st}^{-1}(A)\cap T_S)\overline{\lambda}(\mathrm{d}i). \tag{A.6}$$

By Lemma A.11, we have $Q_{\mathsf{st}(i)}(\mathsf{st}(t),\mathsf{st}(x))(A) = \overline{\mathbb{Q}_i(t,x)}(\mathsf{st}^{-1}(A)\cap T_S)$ for all $i\in T_\Theta$. Thus, by Theorem A.9, we have

$$\int_{T_\Theta}\overline{\mathbb{Q}_i(t,x)}(\mathsf{st}^{-1}(A)\cap T_S)\overline{\lambda}(\mathrm{d}i) = \int_\Theta Q_\theta(\mathsf{st}(t),\mathsf{st}(x))(A)\lambda_p(\mathrm{d}\theta) = Q_{\lambda_p}(\mathsf{st}(t),\mathsf{st}(x))(A). \tag{A.7}$$

Hence, we have the desired result. □

Recall that $\nu \in {}^*\Delta(T_\Theta)$ is the hyperfinite belief function that associates with the Berk-Nash S-equilibrium $m$. We consider the Bellman equation

$$V(s) = \max_{x \in X} \int_S \{\pi(s,x,s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|s,x). \tag{A.8}$$

By the Banach fixed point theorem, there exists an unique $V \in \mathcal{C}[T]$ that is a solution to this Bellman equation. We fix $V$ for the rest of this section.

Similarly, we consider the hyperfinite Bellman equation

$$\mathbb{V}(s) = \max_{x \in T_X} \int_{T_S} \{\Pi(s,x,s') + \delta \mathbb{V}(s')\} \bar{\mathbb{Q}}_{\nu}(\mathrm{d}s'|s,x) \tag{A.9}$$

where $\mathbb{V} : T_S \to \mathbb{R}$ is the unique solution to the hyperfinite Bellman equation. The existence of such $\mathbb{V}$ is guaranteed by the transfer principle. We fix $\mathbb{V}$ for the rest of this section. Define $\mathbb{V}' : {}^*S \to {}^*\mathbb{R}$ by letting $\mathbb{V}'(s) = \mathbb{V}(t_s)$ for all $s \in {}^*T$, where $t_s$ is the unique element in $T_S$ such that $s \in B_S(t_s)$.

**Lemma A.19.** *For all $s \in {}^*S$, $\mathbb{V}'(s) \approx {}^*V(s)$.*

*Proof.* Let $V_0$ be the restriction of ${}^*V$ on $T_S$. For all $(s,x) \in T_S \times T_X$, by Lemma A.18 and Theorem A.9, we have

$$\int_{T_S} \{\Pi(s,x,s') + \delta V_0(s')\} \bar{\mathbb{Q}}_{\nu}(\mathrm{d}s'|s,x) \tag{A.10}$$

$$\approx \int_S \{\pi(\mathsf{st}(s),\mathsf{st}(x),s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s),\mathsf{st}(x)). \tag{A.11}$$

Hence, we have

$$\max_{x \in T_X} \int_{T_S} \{\Pi(s,x,s') + \delta V_0(s')\} \bar{\mathbb{Q}}_{\nu}(\mathrm{d}s'|s,x) \tag{A.12}$$

$$\approx \max_{x \in X} \int_S \{\pi(\mathsf{st}(s),\mathsf{st}(x),s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s),\mathsf{st}(x)) \tag{A.13}$$

$$= V(\mathsf{st}(s)) \approx V_0(s) \tag{A.14}$$

Let $G(f)(s) = \max_{x \in T_X} \int_{T_S} \{\Pi(s,x,s') + \delta f(s')\} \bar{\mathbb{Q}}_{\nu}(\mathrm{d}s'|s,x)$ for all internal function $f : T_S \to {}^*\mathbb{R}$. Note that we have ${}^*d_{\sup}(G(f_1),G(f_2)) \leq \delta {}^*d_{\sup}(f_1,f_2)$ for all internal functions $f_1, f_2 : T_S \to {}^*\mathbb{R}$. Moreover, we can find $\mathbb{V}$ as following: start with $V_0$ and define a sequence $\{V_n\}_{n \in {}^*\mathbb{N}}$ by $V_{n+1} = G(V_n)$. Then $\mathbb{V}$ is the ${}^*$limit of $\{V_n\}_{n \in {}^*\mathbb{N}}$. So:

$${}^*d_{\sup}(V_0,\mathbb{V}) \leq \frac{1}{1-\delta} {}^*d_{\sup}(V_1,V_0) \approx 0. \tag{A.15}$$

As $V$ is continuous, we conclude that $\mathbb{V}'(s) \approx {}^*V(s)$ for all $s \in {}^*S$. □

We now prove the main result of this section.

**Theorem A.20.** *For every $(s, x) \in S \times X$ that is in the support of $m_p$, $x$ is optimal given $s$ in the MDP($\bar{Q}_{\nu_p}$).*

*Proof.* Pick some $(s, x) \in S \times X$ in the support of $m_p$. Then there exists some $(a, b) \in T_S \times T_X$ such that $(a, b) \approx (s, x)$ and $m\big(\{(a, b)\}\big) > 0$. As $m$ is a Berk-Nash S-equilibrium, $b$ is S-optimal given $a$ in HMDP($\bar{\mathbb{Q}}_\nu$). That is, we have

$$\int_{T_S} \{\Pi(a, b, s') + \delta\mathbb{V}(s')\}\bar{\mathbb{Q}}_\nu(\mathrm{d}s'|a, b) \approx \max_{y \in T_X} \int_{T_S} \{\Pi(a, y, s') + \delta\mathbb{V}(s')\}\bar{\mathbb{Q}}_\nu(\mathrm{d}s'|a, y). \tag{A.16}$$

By Lemma A.19, Lemma A.18 and Theorem A.9, we have

$$\int_{T_S} \{\Pi(a, y, s') + \delta\mathbb{V}(s')\}\bar{\mathbb{Q}}_\nu(\mathrm{d}s'|a, y) \approx \int_S \{\pi(s, \mathsf{st}(y), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \mathsf{st}(y)) \tag{A.17}$$

for all $y \in T_X$. Thus, we have $x \in \arg\max_{\hat{x} \in X} \int_S \{\pi(s, \hat{x}, s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \hat{x})$, which implies that $x$ is optimal given $s$ in the MDP($\bar{Q}_{\nu_p}$). □

We now furnish the proof of Theorem 4.11 which concludes the proof of Theorem 1.

***Proof of Theorem 4.11***. Let $m \in {}^*\Delta(T_S \times T_X)$ be a Berk-Nash S-equilibrium with the associated hyperfinite belief $\nu \in {}^*\Delta(T_\Theta)$ for the hyperfinite SMDP. By Theorem A.20, Theorem A.17 and Theorem A.12, $m_p$ is a Berk-Nash equilibrium for the regular SMDP with associated belief function $\nu_p$. □

A.2. **Proofs of Theorems 2 and 3.** For every $n \in \mathbb{N}$ and every finite $\Theta' \subset \hat{\Theta}$, we denote the truncation by $\mathcal{M}^n_{\Theta'}$.

**Lemma A.21.** *Suppose Assumption 1 holds. Then, for every $n \in \mathbb{N}$, the mappings $(s, x) \to Q^n(s, x)$ and $(\theta, s, x) \to Q^n_\theta(s, x)$ are continuous in Prokhorov metric.*

*Proof.* Let $(s_m, x_m)_{m \in \mathbb{N}}$ be a sequence of points in $S_n \times X$ that converges to some point $(s, x) \in S_n \times X$. Let $A$ be a continuity set of $Q^n(s, x)$. As $S_n$ is a continuity set of $Q^n(s, x)$, $A$ is a continuity set of $Q(s, x)$. Thus, we have

$$\lim_{m \to \infty} Q^n(s_m, x_m)(A) = \lim_{m \to \infty} \frac{Q(s_m, x_m)(A)}{Q(s_m, x_m)(S_n)} = \frac{Q(s, x)(A)}{Q(s, x)(S_n)} = Q^n(s, x)(A). \tag{A.18}$$

Thus, the mapping $(s, x) \to Q^n(s, x)$ is continuous in Prokhorov metric. By the same argument, the mapping $(\theta, s, x) \to Q^n_\theta(s, x)$ is continuous in Prokhorov metric. □

**Lemma A.22.** *Suppose Assumption 1 holds. For every $n \in \mathbb{N}$, $Q^n(s,x)$ is dominated by $Q^n_\theta(s,x)$ for all $\theta \in \Theta'$ and all $(s,x) \in S_n \times X$.*

*Proof.* Pick $n \in \mathbb{N}$, $\theta \in \Theta'$ and $(s,x) \in S_n \times X$. Pick some $A \in \mathcal{B}[S_n]$ such that $Q^n_\theta(s,x)(A) = \frac{Q_\theta(s,x)(A)}{Q_\theta(s,x)(S_n)} = 0$. This implies that $Q_\theta(s,x)(A) = 0$. As $\theta \in \Theta' \subset \hat{\Theta}$, we have $Q(s,x)(A) = 0$, which implies that $Q^n(s,x)(A) = 0$ □

For every $n \in \mathbb{N}$, $\theta \in \Theta'$ and every $(s,x) \in S_n \times X$, we use $D_{\theta,n}(\cdot|s,x)$ to denote the density function of $Q^n(s,x)$ with respect to $Q^n_\theta(s,x)$.

**Lemma A.23.** *Suppose Assumption 1 holds. For every $n \in \mathbb{N}$ and $\theta \in \Theta'$, $D_{\theta,n}(s'|s,x)$ is a jointly continuous function of $s'$, $s$ and $x$.*

*Proof.* Pick $n \in \mathbb{N}$ and $\theta \in \Theta'$. For any $A \in \mathcal{B}[S_n]$ and any $(s,x) \in S_n \times X$, we have

$$Q^n(s,x)(A) = \int_A \frac{D_\theta(s'|s,x)}{Q(s,x)(S_n)} Q_\theta(s,x)(\mathrm{d}s') = \int_A \frac{D_\theta(s'|s,x)}{Q(s,x)(S_n)} Q_\theta(s,x)(S_n) Q^n_\theta(s,x)(\mathrm{d}s'). \tag{A.19}$$

So $D_{\theta,n}(s'|s,x) = \frac{D_\theta(s'|s,x)}{Q(s,x)(S_n)} Q_\theta(s,x)(S_n)$. Note that $Q(s,x)(S_n) > 0$ and $Q_\theta(s,x)(S_n) > 0$, and $S_n$ is a continuity set for both $Q(s,x)$ and $Q_\theta(s,x)$. Thus, $D_{\theta,n}(s'|s,x)$ is a jointly continuous function of $s'$, $s$ and $x$. □

As $\Theta'$ is finite, $S_n$ and $X$ are compact, by Lemma A.23, $D_{\theta,n}(s'|s,x)$ is bounded. Hence, Item 5 of Definition 2.4 is automatically satisfied for the SMDP $\mathcal{M}^n_{\Theta'}$. Moreover, the payoff function $\pi_n$ is continuous on $S_n \times X \times S_n$. Hence, by Theorem 1, we have

**Lemma A.24.** *Suppose Assumption 1 holds. For every $n \in \mathbb{N}$ and every finite $\Theta' \subset \hat{\Theta}$, the SMDP $\mathcal{M}^n_{\Theta'}$ is regular and has a Berk-Nash equilibrium.*

Theorem 4.12 then follows from the transfer of Lemma A.24. Let $m \in {}^*\Delta({}^*S_N \times {}^*X)$ denote the Berk-Nash *equilibrium of $\mathcal{M}^N_{T_\Theta}$, with the associated *belief $\nu$. Assumption 2 guarantees that the push-down, $m_p$, of $m$ is a probability measure on $\Delta(S \times X)$. To show that $m_p$ is a Berk-Nash equilibrium for the original SMDP $\mathcal{M}$ with the associated belief function $\nu_p$, we break the proof into three subsections (A.2.1-A.2.5) which establishes stationarity, optimality and belief restriction, respectively.

A.2.1. ***Stationarity***. In this section, we show that $m_p$ satisfies stationarity. Using essentially the same argument as in Lemma A.8, we have the following result.

**Lemma A.25.** *For all $A \in \mathcal{B}[S]$, $(m_p)_S(A) = \int_{{}^*S \times {}^*X} \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))\overline{m}(\mathrm{d}s, \mathrm{d}x)$, where $(m_p)_S$ denote the marginal measure of $m_p$ on $S$.*

**Lemma A.26.** *For every* $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$*, every* $\theta \in T_\Theta$ *and every* $A \in \mathcal{B}[S]$*, we have* $Q(\mathsf{st}(s), \mathsf{st}(x))(A) = \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))$ *and* $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))(A) = \overline{{}^*Q^N_\theta(s,x)}(\mathsf{st}^{-1}(A))$*.*

*Proof.* Pick some $(s_0, x_0) \in \mathrm{NS}({}^*S) \times {}^*X$, some $\theta_0 \in T_\Theta$ and some $A_0 \in \mathcal{B}[S]$. By Theorem A.10, we have $Q(\mathsf{st}(s_0), \mathsf{st}(x_0))(A_0) = \overline{{}^*Q(s_0,x_0)}(\mathsf{st}^{-1}(A_0))$. As ${}^*Q^N(s_0,x_0)({}^*S_N) \approx 1$, we have $Q(\mathsf{st}(s_0), \mathsf{st}(x_0))(A_0) = \overline{{}^*Q^N(s_0,x_0)}(\mathsf{st}^{-1}(A_0))$. By the same argument, we have $Q_{\mathsf{st}(\theta_0)}(\mathsf{st}(s_0), \mathsf{st}(x_0))(A_0) = \overline{{}^*Q^N_{\theta_0}(s_0,x_0)}(\mathsf{st}^{-1}(A_0))$. □

**Theorem A.27.** *Suppose Assumption 1 holds. For every* $A \in \mathcal{B}[S]$*,* $(m_p)_S(A) = \int_{S\times X} Q(A|s,x) m_p(\mathrm{d}s, \mathrm{d}x)$*.*

*Proof.* Pick $A \in \mathcal{B}[S]$. By Lemma A.25, $(m_p)_S(A) = \int_{{}^*S\times{}^*X} \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))\overline{m}(\mathrm{d}s,\mathrm{d}x)$ As $\overline{m}(\mathsf{st}^{-1}(S) \times {}^*X) = 1$, we have

$$\int_{{}^*S\times{}^*X} \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))\overline{m}(\mathrm{d}s,\mathrm{d}x) = \lim_{n\to\infty} \int_{{}^*S_n\times{}^*X} \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))\overline{m}(\mathrm{d}s,\mathrm{d}x). \tag{A.20}$$

By Lemma A.26 and Theorem A.9, we have

$$\int_{{}^*S_n\times{}^*X} \overline{{}^*Q^N(s,x)}(\mathsf{st}^{-1}(A))\overline{m}(\mathrm{d}s,\mathrm{d}x) = \int_{S_n\times X} Q(s,x)(A) m_p(\mathrm{d}s,\mathrm{d}x). \tag{A.21}$$

Note that we also have

$$\lim_{n\to\infty} \int_{S_n\times X} Q(s,x)(A) m_p(\mathrm{d}s,\mathrm{d}x) = \int_{S\times X} Q(s,x)(A) m_p(\mathrm{d}s,\mathrm{d}x). \tag{A.22}$$

So, we have the desired result. □

A.2.2. ***Belief Restriction under Assumption 3.*** In this section, we establish belief restriction assuming uniformly bounded relative entropy. Recall that $\nu$ is the hyperfinite belief that associates with the Berk-Nash ${}^*$equilibrium $m$ of the nonstandard SMDP $\mathcal{M}^N_{T_\Theta}$. Recall that $T_\Theta \subset {}^*\hat{\Theta}$. By the transfer of Lemma A.22, ${}^*Q^N(s,x)$ is ${}^*$dominated by ${}^*Q^N_\theta(s,x)$ for all $\theta \in T_\Theta$ and $(s,x) \in {}^*S_N \times {}^*X$. We use $\mathbb{D}_\theta(\cdot|s,x)$ to denote the ${}^*$density function of ${}^*Q^N(s,x)$ with respect to ${}^*Q^N_\theta(s,x)$. By the transfer of Lemma A.23, $\mathbb{D}_\theta(s'|s,x)$ is jointly ${}^*$continuous on $(s',s,x)$.

**Lemma A.28.** *Suppose Assumption 1 holds. For all* $\theta \in T_\Theta$ *and all* $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$*, we have* $\mathbb{D}_\theta(s'|s,x) \approx {}^*D_\theta(s'|s,x)$ *on a* ${}^*Q^N_\theta(s,x)$ *measure* 1 *set.*

*Proof.* Pick $\theta \in T_\Theta$ and $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$. Note that ${}^*Q(s,x)({}^*S_N) \approx 1$ and ${}^*Q_\theta(s,x)({}^*S_N) \approx 1$. Thus, for every $A \in {}^*\mathcal{B}[{}^*S_N]$, we have:

$$ {}^*Q^N(s,x)(A) = \frac{{}^*Q(s,x)(A)}{{}^*Q(s,x)({}^*S_N)} \approx \int_{{}^*S_N} {}^*D_\theta(s'|s,x){}^*Q_\theta^N(s,x)(\mathrm{d}s'). \tag{A.23} $$

Note that ${}^*Q^N(s,x)(A) = \int_{{}^*S_N} \mathbb{D}_\theta(s'|s,x){}^*Q_\theta^N(s,x)(\mathrm{d}s')$. Thus, we conclude that $\mathbb{D}_\theta(s'|s,x) \approx {}^*D_\theta(s'|s,x)$ on some ${}^*Q_\theta^N(s,x)$ measure 1 set. □

For every $\theta \in T_\Theta$, let the nonstandard Kullback-Leibler divergence be:

$$ {}^*K_N(m,\theta) = \int_{{}^*S_N \times {}^*X} {}^*\mathbb{E}_{{}^*Q^N(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right] m(\mathrm{d}s,\mathrm{d}x). \tag{A.24} $$

The set of closest parameter values given $m$ is the set $T_\Theta^N(m) = \arg\min_{\theta \in T_\Theta} {}^*K_N(m,\theta)$. Recall that we use $\Theta_{m_p}$ to denote the set $\{\theta \in \Theta : K_Q(\theta, m_p) < \infty\}$. The proof of the following lemma is provided in Appendix C.1.

**Lemma A.29.** *Suppose Assumption 1 and Assumption 2 hold. For every* $(\theta,s,x) \in T_\Theta \times \mathrm{NS}({}^*S) \times {}^*X$*, if* $\mathsf{st}(\theta) \in \Theta_{m_p}$ *and* $Q(\mathsf{st}(s),\mathsf{st}(x))$ *is dominated by* $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s),\mathsf{st}(x))$*, then* ${}^*\mathbb{E}_{{}^*Q^N(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right] \approx \mathbb{E}_{Q(\cdot|\mathsf{st}(s),\mathsf{st}(x))}\left[\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right)\right]$.

By the transfer principle, ${}^*Q^N(s,x)({}^*S_N) > r$ for all $(s,x) \in {}^*S_N \times {}^*X$. As $T_\Theta \subset {}^*\hat{\Theta}$, following the calculation in Lemma A.23, $|\mathbb{D}_\theta(s'|s,x)| \leq \frac{1}{r}|{}^*D_\theta(s'|s,x)|$ for all $(s',s,\theta,x) \in {}^*S_N \times {}^*S_N \times T_\Theta \times {}^*X$. We now establish the connections between the nonstandard weighted Kullback-Leibler divergence ${}^*K_N(m,\theta)$ and the standard weighted Kullback-Leibler divergence $K_Q(m_p,\theta)$.

**Theorem A.30.** *Suppose Assumption 1, Assumption 2 and Assumption 3 hold. For every* $\theta \in T_\Theta$*, if* $\mathsf{st}(\theta) \in \Theta_{m_p}$*, then* ${}^*K_N(m,\theta) \approx K_Q(m_p,\mathsf{st}(\theta))$.

*Proof.* Pick $\theta \in T_\Theta$ such that $\mathsf{st}(\theta) \in \Theta_{m_p}$. Since $K_Q(m_p,\mathsf{st}(\theta)) < \infty$, $Q(s,x)$ is dominated by $Q_{\mathsf{st}(\theta)}(s,x)$ for $m_p$-almost all $(s,x) \in S \times X$. Let $\bar{\mathbb{R}}$ denote the extended real line and define $g : S \times X \to \bar{\mathbb{R}}$ to be $g(s,x) = \mathbb{E}_{Q(\cdot|s,x)}\left[\ln\left(D_{\mathsf{st}(\theta)}(s'|s,x)\right)\right]$ if $Q(s,x)$ is dominated by $Q_{\mathsf{st}(\theta)}(s,x)$ and $g(s,x) = \infty$ otherwise. We have

$$ K_Q(m_p,\mathsf{st}(\theta)) = \int_{S \times X} g(s,x)m_p(\mathrm{d}s,\mathrm{d}x) = \lim_{n\to\infty} \int_{S_n \times X} g(s,x)m_p(\mathrm{d}s,\mathrm{d}x). \tag{A.25} $$

Let $G : {}^*S_N \times {}^*X \to {}^*\mathbb{R}$ be $G(s,x) = {}^*\mathbb{E}_{{}^*Q^N(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right]$. By Lemma A.29 and Theorem A.9, we have $\int_{S_n \times X} g(s,x)m_p(\mathrm{d}s,\mathrm{d}x) \approx \int_{S_n \times X} G(s,x)m(\mathrm{d}s,\mathrm{d}x)$. To finish the proof, it is sufficient to show that $G$ is S-integrable with respect to $m$. As

$\theta \in T_\Theta \subset {}^*\hat{\Theta}$ and $m$ is *stationary, by Assumption 3, ${}^*\mathbb{E}_{{}^*Q(\cdot|s,x)}\left[\ln\left({}^*D_\theta(s'|s,x)\right)\right]$ is S-integrable with respect to $m$. By Item 3 of Assumption 1, ${}^*\mathbb{E}_{{}^*Q^N(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right]$ is S-integrable with respect $m$, completing the proof. □

We now prove the main result of this section.

**Theorem A.31.** *Suppose Assumption 1, Assumption 2 and Assumption 3 hold. The support of $\nu_p$ is a subset of $\Theta_Q(m_p)$.*

*Proof.* Pick $\theta_0 \in \Theta$ such that $\theta_0$ is in the support of $\nu_p$. As $\nu_p(A) = \overline{\nu}(\mathsf{st}^{-1}(A))$ for all $A \in \mathcal{B}[\Theta]$, by Theorem 4.9, there exists $\theta_1 \approx \theta_0$ such that $\theta_1 \in T_\Theta^N(m)$. That is, we have ${}^*K_N(m,\theta_1) = \arg\min_{t\in T_\Theta} {}^*K_N(m,t)$. Suppose there exists $\theta' \in \Theta$ such that $K_Q(m_p,\theta') < K_Q(m_p,\theta_0) - \frac{1}{n}$ for some $n \in \mathbb{N}$. Clearly, both $\theta'$ and $\theta_0$ belong to $\Theta_{m_p}$. Let $t_{\theta'} \in T_\Theta$ be the unique element such that $\theta' \in B_\Theta(t_{\theta'})$. By Theorem A.30, we have

$$ {}^*K_N(m,t_{\theta'}) \approx K_Q(m_p,\theta') < K_Q(m_p,\theta_0) - \frac{1}{n} \lessapprox {}^*K_N(m,\theta_1) - \frac{1}{n}. \tag{A.26}$$

This is a contradiction, hence we conclude that $\nu_p$ is a subset of $\Theta_Q(m_p)$. □

A.2.3. ***Belief Restriction without Assumption 3***. In this section, we establish belief restriction of the SMDP $\mathcal{M}$ if $\mathcal{M}$ is correctly specified or satisfies Assumption 4. We first assume that $\mathcal{M}$ is correctly specified.

**Theorem A.32.** *Suppose the SMDP $\mathcal{M}$ is correctly specified, Assumption 1 and Assumption 2 hold. Then, the support of $\nu_p$ is a subset of $\Theta_Q(m_p)$.*

*Proof.* As the SMDP $\mathcal{M}$ is correctly specified and $\Theta \subset T_\Theta$, we have $\min_{t\in T_\Theta} {}^*K_N(m,t) = 0$. Pick $\theta_0 \in \Theta$ and $(s_0,x_0)$ such that $\theta_0$ in the support of $\nu_p$ and $(s_0,x_0)$ in the support of $m_p$. Then, by Theorem 4.9, there exist $\theta_1 \approx \theta_0$ and $(s_1,x_1) \in {}^*S_N \times {}^*X$ such that $\theta_1 \in T_\Theta^N(m)$ and $(s_1,x_1)$ in the *support of $m$. By the transfer of Lemma 1 in EP, we have ${}^*Q^N_{\theta_1}(s_1,x_1) = {}^*Q^N(s_1,x_1)$. As $s_1$ is near-standard, by Definition 2.4, we conclude that $Q_{\theta_0}(s_0,x_0) = Q(S_0,x_0)$ and therefore, $K_Q(m_p,\theta_0) = 0$. □

We now assume that Assumption 4 holds but $\mathcal{M}$ may be misspecified.

**Theorem A.33.** *Suppose Assumption 1, Assumption 2 and Assumption 4 hold. Then the support of $\nu_p$ is a subset of $\Theta_Q(m_p)$.*

*Proof.* As $\Theta \subset T_\Theta$, by Assumption 2, $T_\Theta^N(m) = \{\theta_0\}$. Thus, we have $\nu_p(\{\theta_0\}) = \nu(\{\theta_0\}) = 1$. By Assumption 4 again, $\Theta_Q(m_p) = \{\theta_0\}$, completing the proof. □

A.2.4. ***Optimality with Bounded Payoff Function.*** In this appendix, we establish optimality of the candidate Berk-Nash equilibrium $m_p$ assuming bounded and continuous payoff function. We start with the follwing lemma:

**Lemma A.34.** *For every $\lambda \in {}^*\Delta(T_\Theta)$ and every $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$, $({}^*\bar{Q}^N_\lambda(s,x))_p = \bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))$. That is, the push-down of ${}^*\bar{Q}^N_\lambda(s,x)$ is the same as $\bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))$.*

*Proof.* Fix $(\lambda, s, x) \in {}^*\Delta(T_\Theta) \times \mathrm{NS}({}^*S) \times {}^*X$ and $A \in \mathcal{B}[S]$. As $\overline{{}^*\bar{Q}^N_\lambda(s,x)}(\mathsf{st}^{-1}(S)) = 1$, we have $({}^*\bar{Q}^N_\lambda(s,x))_p(A) = \overline{{}^*\bar{Q}_\lambda(s,x)}(\mathsf{st}^{-1}(A)) = \int_{T_\Theta} \overline{{}^*Q_i(s,x)}(\mathsf{st}^{-1}(A))\overline{\lambda}(\mathrm{d}i)$. By Lemma A.11, we have $Q_{\mathsf{st}(i)}(\mathsf{st}(s), \mathsf{st}(x))(A) = \overline{{}^*Q_i(s,x)}(\mathsf{st}^{-1}(A))$ for all $i \in T_\Theta$. Thus, by Theorem A.9, we have

$$\int_{T_\Theta} \overline{{}^*Q_i(s,x)}(\mathsf{st}^{-1}(A))\overline{\lambda}(\mathrm{d}i) = \int_\Theta Q_\theta(\mathsf{st}(s), \mathsf{st}(x))(A)\lambda_p(\mathrm{d}\theta) = \bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))(A). \tag{A.27}$$

Hence, we have the desired result. □

We now consider the Bellman equation

$$V(s) = \max_{x \in X} \int_S \{\pi(s,x,s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s,x). \tag{A.28}$$

Let $\mathcal{C}_0[S]$ denote the set of bounded continuous functions on $S$ equipped with the sup-norm. Then $\mathcal{C}_0[S]$ is a complete metric space. Under Assumption 5, the map $F(g)(s) = \max_{x \in X} \int_S \{\pi(s,x,s') + \delta g(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s,x)$ is a contraction mapping from $\mathcal{C}_0[S]$ to $\mathcal{C}_0[S]$. By the Banach fixed point theorem, there is an unique $V \in \mathcal{C}_0[S]$ that is a solution to the Bellman equation Eq. (A.28). We fix $V$ for the rest of this section. Similarly, we consider the nonstandard Bellman equation

$$\mathbb{V}(s) = \max_{x \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta \mathbb{V}(s')\}{}^*\bar{Q}^N_\nu(\mathrm{d}s'|s,x). \tag{A.29}$$

where $\mathbb{V} \in {}^*\mathcal{C}_0[{}^*S_N]$ is the unique solution of the nonstandard Bellman equation Eq. (A.29). The existence of such $\mathbb{V}$ is guaranteed by the transfer principle. We also fix $\mathbb{V}$ for the rest of this section.

**Lemma A.35.** *Suppose Assumption 1 and Assumption 5 hold. For every $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$:*

$$\int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*V(s')\}{}^*\bar{Q}^N_\nu(\mathrm{d}s'|s,x) \tag{A.30}$$

$$\approx \int_S \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s), \mathsf{st}(x)). \tag{A.31}$$

*Proof.* Pick $(t,x) \in \mathrm{NS}({}^*T) \times {}^*X$. As $\pi$ is bounded, ${}^*\pi_N(t,x,\cdot)$ is bounded. By Arkeryd et al. (1997, Section 4, Corollary 6.1), ${}^*\pi_N(t,x,t') + \delta {}^*V(t')$ is S-integrable with respect to ${}^*\bar{Q}_{\nu}^{N}(\mathrm{d}t'|t,x)$. By Arkeryd et al. (1997, Section 4, Corollary 6.1), Lemma A.18 and Theorem A.9 , we have

$$\int_{{}^*T_N} \{{}^*\pi_N(t,x,t') + \delta {}^*V(t')\} {}^*\bar{Q}_{\nu}^{N}(\mathrm{d}t'|t,x) \tag{A.32}$$

$$= \lim_{n\to\infty} \int_{T_n} \{\pi(\mathsf{st}(t),\mathsf{st}(x),t') + \delta V(t')\} \bar{Q}_{\nu_p}(\mathrm{d}t'|\mathsf{st}(t),\mathsf{st}(x)) \tag{A.33}$$

$$= \int_{T} \{\pi(\mathsf{st}(t),\mathsf{st}(x),t') + \delta V(t')\} \bar{Q}_{\nu_p}(\mathrm{d}t'|\mathsf{st}(t),\mathsf{st}(x)). \tag{A.34}$$

Hence, we have the desired result. □

The set $\mathcal{C}_0[S]$ is a complete metric space under the metric $d_{\sup}$. Recall that, under Assumption 1, $S_n$ is a non-decreasing sequence of compact subsets of $S$ such that $S = \bigcup_{n\in\mathbb{N}} S_n$. For two elements $g_1, g_2 \in \mathcal{C}_0[S]$, define $d_{\sup,n}(g_1,g_2) = \sup_{s\in S_n} |g_1(s) - g_2(s)|$. Define $d_{\mathrm{unif}}(g_1,g_2) = \sum_{n\in\mathbb{N}} \frac{\min\{1, d_{\sup,n}(g_1,g_2)\}}{2^n}$. Note that $d_{\mathrm{unif}}$ is a well-defined complete metric on $\mathcal{C}_0[S]$. For every $f \in \mathcal{C}_0[S]$, under the topology generated by the metric $d_{\mathrm{unif}}$, $F \in {}^*\mathcal{C}_0[{}^*S]$ is in the monad of ${}^*f$ if $F(s) \approx {}^*f(s)$ for all $s \in \mathrm{NS}({}^*S)$.

**Lemma A.36.** *Suppose Assumption 1 and Assumption 5 hold. Then* ${}^*V(s) \approx \mathbb{V}(s)$ *for all* $s \in \mathrm{NS}({}^*S)$.

*Proof.* Let $V_0$ be the restriction of ${}^*V$ to ${}^*S_N$. For all $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$, by Lemma A.35, we have

$$\max_{x\in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta V_0(s')\} {}^*\bar{Q}_{\nu}^{N}(\mathrm{d}s'|s,x) = V(\mathsf{st}(s)) \approx V_0(s). \tag{A.35}$$

Let $G(f)(s) = \max_{x\in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*f(s')\} {}^*\bar{Q}_{\nu}^{N}(\mathrm{d}s'|s,x)$ for all $f \in {}^*\mathcal{C}_0({}^*S_N)$. Consider the following internal iterated process: start with $V_0$ and define a sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ by $V_{n+1} = G(V_n)$. As $\delta \in [0,1)$ and ${}^*S_N$ is a ${}^*$compact set, there exists some $K \in {}^*\mathbb{N}$ such that ${}^*d_{\sup}(V_K, V_{K+1}) < 1$. Hence the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ is a ${}^*$Cauchy sequence with respect to the ${}^*$metric ${}^*d_{\mathrm{unif}}$. As ${}^*\mathcal{C}_0({}^*S_N)$ is ${}^*$complete with respect to ${}^*d_{\mathrm{unif}}$, the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ has a ${}^*$limit. Note that ${}^*d_{\mathrm{unif}}(G(f_1), G(f_2)) \leq {}^*d_{\mathrm{unif}}(f_1,f_2)$ for all $f_1, f_2 \in {}^*\mathcal{C}_0({}^*S_N)$. So $G$ is a ${}^*$continuous function, hence the ${}^*$limit of the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ is the ${}^*$fixed point $\mathbb{V}$. As ${}^*d_{\mathrm{unif}}(V_0, \mathbb{V}) \approx {}^*d_{\mathrm{unif}}(V_1, V_0) \approx 0$, we have ${}^*V(s) \approx \mathbb{V}(s)$ for all $s \in \mathrm{NS}({}^*S)$. □

**Theorem A.37.** *Suppose Assumption 1 and Assumption 5 hold. For every* $(s,x) \in S \times X$ *that is in the support of* $m_p$*,* $x$ *is optimal given* $s$ *in the MDP(*$\bar{Q}_{\nu_p}$*).*

*Proof.* Pick $(s,x) \in S \times X$ that is in the support of $m_p$. Then there exists some $(a,b) \in \mathrm{NS}({}^*S) \times {}^*X$ such that $(a,b) \approx (s,x)$ and $(a,b)$ is in the ${}^*$support of $m$. Thus, we have $b \in \arg\max_{y \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(a,y,s') + \delta\mathbb{V}(s')\}{}^*\bar{Q}_\nu^N(\mathrm{d}s'|a,y)$.

**Claim A.38.** *$\mathbb{V}$ is bounded.*

*Proof of Claim A.38.* Let $G(f)(s) = \max_{x \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*f(s')\}{}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x)$ for all $f \in {}^*\mathcal{C}_0({}^*S_N)$ and $F_0 : {}^*S_N \to {}^*\mathbb{R}$ be the constant 0 function. Consider the following internal iterated process: start with $F_0$ and define a sequence $\{F_n\}_{n \in {}^*\mathbb{N}}$ by $F_{n+1} = G(F_n)$. The ${}^*$limit (with respect to ${}^*d_{\sup}$) of the internal sequence $\{F_n\}_{n \in {}^*\mathbb{N}}$ is $\mathbb{V}$. By the transfer of the Banach fixed point theorem, we know that ${}^*d_{\sup}(F_0, \mathbb{V}) \leq \frac{1}{1-\delta} {}^*d_{\sup}(F_0, F_1)$. As ${}^*\pi_N$ is bounded, we conclude that $\mathbb{V}$ is bounded.

By Claim A.38, Arkeryd et al. (1997, Section 4, Corollary 6.1), Lemma A.36, Lemma A.34 and Theorem A.9:

$$\int_{{}^*S_N} \{{}^*\pi_N(a,y,s') + \delta\mathbb{V}(s')\}{}^*\bar{Q}_\nu^N(\mathrm{d}s'|a,y) \tag{A.36}$$

$$\approx \lim_{n\to\infty} \int_{S_n} \{\pi(s, \mathsf{st}(y), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \mathsf{st}(y)) \tag{A.37}$$

$$= \int_S \{\pi(s, \mathsf{st}(y), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \mathsf{st}(y)). \tag{A.38}$$

for all $y \in {}^*X$. Thus, we have $x \in \arg\max_{\hat{x} \in X} \int_S \{\pi(s,\hat{x},s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s,\hat{x})$, which implies that $x$ is optimal given $s$ in the MDP($\bar{Q}_{\nu_p}$). ☐

A.2.5. ***Optimality with Unbounded Payoff Function***. In this appendix, we establish optimality of of the candidate Berk-Nash equilibrium $m_p$ with possibly unbounded payoff function under Assumption 6, Assumption 7 and Assumption 8. Let $\|s\|, d_S$ denote the norm of an element $s \in S$ and the metric on $S$, respectively. Let $W(\mu,\nu)$ denote the Wasserstein distance between two probability measures $\mu$ and $\nu$.

**Lemma A.39.** *Suppose Assumption 8 holds. For every $\lambda \in {}^*\Delta(T_\Theta)$ and every $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$, ${}^*W\big({}^*\bar{Q}_\lambda(s,x), {}^*\bar{Q}_{{}^*\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))\big) \approx 0$. That is, ${}^*\bar{Q}_\lambda(s,x)$ is in the monad of $\bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))$ with respect to the 1-Wasserstein metric.*

*Proof.* Fix $\lambda \in {}^*\Delta(T_\Theta)$ and $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$. Note that convergence in the Wasserstein metric is equivalent to weak convergence plus convergence of the first moments. By Lemma A.34, it is sufficient to show that

$$\int_{{}^*S} {}^*d_S(t, s_0){}^*\bar{Q}_\lambda(s,x)(\mathrm{d}t) \approx \int_S d_S(t, \mathsf{st}(s_0))\bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))(\mathrm{d}t) \tag{A.39}$$

for all $s_0 \in \mathrm{NS}({}^*S)$. By Assumption 8, we have

$$\int_{{}^*S} {}^*d_S(t, s_0){}^*Q_\theta(s, x)(\mathrm{d}t) \approx \int_S d_S(t, \mathsf{st}(s_0))Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), ST(x))(\mathrm{d}t). \tag{A.40}$$

for all $\theta \in T_\Theta$. By Theorem A.9, we have

$$\int_{{}^*S} {}^*d_S(t, s_0){}^*\bar{Q}_\lambda(s, x)(\mathrm{d}t) \approx \int_\Theta \int_S d_S(t, \mathsf{st}(s_0))Q_\theta(\mathsf{st}(s), \mathsf{st}(x))(\mathrm{d}t)\lambda_p(\mathrm{d}\theta) \tag{A.41}$$

$$= \int_S d_S(t, \mathsf{st}(s_0))\bar{Q}_{\lambda_p}(\mathsf{st}(s), \mathsf{st}(x))(\mathrm{d}t). \tag{A.42}$$

Hence, we have the desired result. $\square$

We now consider the Bellman equation.

$$V(s) = \max_{x \in X} \int_S \{\pi(s, x, s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, x). \tag{A.43}$$

For each $n \in \mathbb{N}$ and any two elements $g_1, g_2 \in \mathcal{C}[S]$ (the set of continuous real-valued functions on $S$), let $d_{\sup,n}(g_1, g_2) = \sup_{s \in S_n} |g_1(s) - g_2(s)|$. Recall that the uniform convergence topology on compact sets on $\mathcal{C}[S]$ can be generated from the metric $d_{\mathrm{unif}}(g_1, g_2) = \sum_{n \in \mathbb{N}} \frac{\min\{1, d_{\sup,n}(g_1, g_2)\}}{2^n}$. Note that $\mathcal{C}[S]$ equipped with $d_{\mathrm{unif}}$ is a complete metric space. Let $B, D$ be constants in Assumption 6 and Assumption 7, respectively. Define $\mathcal{L}_{B,D}[S] = \{f \in \mathcal{C}[S] : (\exists E \in \mathbb{R}_{>0})(\forall s \in S)(|f(s)| \leq E + (B + D)\|s\|)\}$. Then $\mathcal{L}_{B,D}[S]$ is a complete metric space under the metric $d_{\mathrm{unif}}$. We present three lemmas, Lemma A.40- Lemma A.42, proofs of which are provided in the Online Appendix.

**Lemma A.40.** *Suppose Assumption 1, Assumption 6, Assumption 8 and Assumption 7 hold. The Bellman operator* $F(g)(s) = \max_{x \in X} \int_S \{\pi(s, x, s') + \delta g(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, x)$ *maps every element in* $\mathcal{L}_{B,D}[S]$ *to some element in* $\mathcal{L}_{B,D}[S]$.

The Bellman operator $F(g)(s) = \max_{x \in X} \int_S \{\pi(s, x, s') + \delta g(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, x)$ is a contraction mapping on $\mathcal{L}_{B,D}[S]$. Given any $g_0 \in \mathcal{L}_{B,D}[S]$, let $\{g_n\}_{n \geq 0}$ be the sequence such that $g_{n+1} = F(g_n)$ for all $n \geq 0$. The sequence $\{g_n\}_{n \geq 0} \subset \mathcal{L}_{B,D}[S]$ is a Cauchy sequence with respect to the metric $d_{\mathrm{unif}}$. This is because, for every $n \in \mathbb{N}$, there exists some $K \in \mathbb{N}$ such that $d_{\sup,n}(g_K, g_{K+1}) < 1$. The limit of the sequence $\{g_n\}_{n \geq 0}$ is the unique fixed point of the Bellman operator, hence is the solution of the Bellman equation. We use $V$ to denote the unique solution of the Bellman equation Eq. (A.43), and fix this notation for the rest of this section.

**Lemma A.41.** *Suppose Assumption 1, Assumption 6, Assumption 7 and Assumption 8 hold. For every* $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$*:*

$$\int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*V(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x) \tag{A.44}$$

$$\approx \int_S \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s), \mathsf{st}(x)). \tag{A.45}$$

The nonstandard Bellman equation is:

$$\mathbb{V}(t) = \max_{x \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta \mathbb{V}(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x). \tag{A.46}$$

Let ${}^*\mathcal{C}_0[{}^*S_N]$ is the set of ${}^*$bounded continuous functions on ${}^*S_N$. Note that ${}^*\pi_N$ is an element in ${}^*\mathcal{C}_0[{}^*S_N]$. By the transfer of the Banach fixed point theorem, there exists a unique solution $\mathbb{V}$ of the nonstandard Bellman equation Eq. (A.46). We fix $\mathbb{V}$ for the rest of this section.

**Lemma A.42.** *Suppose Assumption 1, Assumption 6, Assumption 7 and Assumption 8 hold. Then* ${}^*V(s) \approx \mathbb{V}(s)$ *for all* $s \in \mathrm{NS}({}^*S)$.

To complete the proof of the main result of this section, we need to show that the solution $\mathbb{V}$ of the nonstandard Bellman equation (Eq. (A.46)) is S-integrable. Let

$${}^*\mathcal{L}_{B,D}[{}^*S_N] = \{f \in {}^*\mathcal{C}_0[{}^*S_N] : (\exists E \in {}^*\mathbb{R}_{>0})(\forall s \in {}^*S_N)(|f(s)| \leq E + (B+D)\|s\|)\}. \tag{A.47}$$

Note that ${}^*\mathcal{L}_{B,D}[{}^*S_N]$ is a ${}^*$closed subset of ${}^*\mathcal{C}_0[{}^*S_N]$ under ${}^*d_{\sup}$, hence is a ${}^*$complete metric space under the ${}^*$metric ${}^*d_{\sup}$.

**Lemma A.43.** *Suppose Assumption 1, Assumption 6 and Assumption 7 hold. The nonstandard Bellman operator, defined below, maps every element in* ${}^*\mathcal{L}_{B,D}[{}^*S_N]$ *to some element in* ${}^*\mathcal{L}_{B,D}[{}^*S_N]$.

$$G(f)(s) = \max_{x \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*f(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x) \tag{A.48}$$

*Proof.* Let $f$ be an arbitrary element in ${}^*\mathcal{L}_{B,D}({}^*S_N)$. Then, there is some $E \in {}^*\mathbb{R}_{>0}$ such that $|f(s)| \leq E + (B+D)\|s\|$ for all $s \in {}^*S_N$. By Assumption 6 and Assumption 7:

$$\left|\int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta^* f(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x)\right| \tag{A.49}$$

$$\leq A + \delta E + \int_{{}^*S} \{B \max\{\|s\|, \|s'\|\} + \delta(B+D)\|s'\|\} {}^*\bar{Q}_\nu(\mathrm{d}s'|s,x) \tag{A.50}$$

$$\leq A + \delta E + \big(B + \delta(B+D)\big)C + (B+D)\|s\|. \tag{A.51}$$

Thus, we have the desired result. □

Hence, we conclude that the solution $\mathbb{V}$ of the the nonstandard Bellman equation (Eq. (A.46)) is an element of ${}^*\mathcal{L}_{B,D}[{}^*S_N]$.

**Lemma A.44.** *Suppose Assumption 1, Assumption 6, Assumption 7 and Assumption 8 hold. Then* $\mathbb{V}$ *is S-integrable with respect to* ${}^*\bar{Q}_\nu^N(s,x)$ *when* $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$.

*Proof.* Pick some $(s,x) \in \mathrm{NS}({}^*S) \times {}^*X$. As $\mathbb{V} \in {}^*\mathcal{L}_{B,D}[{}^*S_N]$, there exist some $E \in {}^*\mathbb{R}$ such that $|\mathbb{V}(t)| \leq E + (B+D)\|t\|$ for all $t \in {}^*S_N$. By Lemma A.42, ${}^*V(t) \approx \mathbb{V}(t)$ for all $t \in \mathrm{NS}({}^*S)$. Hence, $E$ is near-standard. As $D \in \mathbb{R}_{>0}$, it is sufficient to show that $\|t\|$ is S-integrable with respect to ${}^*\bar{Q}_\nu(s,x)$. By Lemma A.39 and Theorem A.9:

$$\int_{{}^*S_N} \|t\| {}^*\bar{Q}_\nu^N(\mathrm{d}t|s,x) \lessapprox \int_{{}^*S} \|t\| {}^*\bar{Q}_\nu(\mathrm{d}t|s,x) \tag{A.52}$$

$$\approx \lim_{n\to\infty} \int_{S_n} \|t\| \bar{Q}_{\nu_p}(\mathrm{d}t|\mathsf{st}(s),\mathsf{st}(x)) \tag{A.53}$$

$$= \lim_{n\to\infty} \mathsf{st}\big(\int_{{}^*S_n} \|t\| {}^*\bar{Q}_\nu^N(\mathrm{d}t|s,x)\big). \tag{A.54}$$

Note that $\int_{{}^*S_N} \|t\| {}^*\bar{Q}_\nu^N(\mathrm{d}t|s,x) \gtrapprox \lim_{n\to\infty} \mathsf{st}\big(\int_{{}^*S_n} \|t\| {}^*\bar{Q}_\nu^N(\mathrm{d}t|s,x)\big)$. Hence, by Arkeryd et al. (1997, Section 4, Theorem 6.2), $\|t\|$ is S-integrable with respect to ${}^*\bar{Q}_\nu^N(s,x)$, completing the proof. □

**Theorem A.45.** *Suppose Assumption 1, Assumption 6, Assumption 7 and Assumption 8 hold. Then, for every* $(s,x) \in S \times X$ *that is in the support of* $m_p$*,* $x$ *is optimal given* $s$ *in the MDP(*$\bar{Q}_{\nu_p}$*).*

*Proof.* Pick $(s,x) \in S \times X$ that is in the support of $m_p$. Then there exists some $(a,b) \in \mathrm{NS}({}^*S) \times {}^*X$ such that $(a,b) \approx (s,x)$ and $m(\{(a,b)\}) > 0$. Thus, we have

$$b \in \arg\max_{y \in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(a,y,s') + \delta \mathbb{V}(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|a,y). \tag{A.55}$$

By Lemma A.44, $\mathbb{V}$ is S-integrable with respect to ${}^*\bar{Q}_\nu^N(\mathrm{d}s'|a,y)$ for all $y \in {}^*X$. Using similar argument, ${}^*\pi_N(a,y,\cdot)$ is also S-integrable with respect to ${}^*\bar{Q}_\nu^N(\mathrm{d}s'|a,y)$ for all $y \in {}^*X$. Thus, by Arkeryd et al. (1997, Section 4, Theorem 6.2), Lemma A.42, Lemma A.34 and Theorem A.9:

$$\int_{{}^*S_N} \{{}^*\pi_N(a, y, s') + \delta\mathbb{V}(s')\}{}^*\bar{Q}^N_\nu(\mathrm{d}s'|a, y) \tag{A.56}$$

$$\approx \lim_{n\to\infty} \int_{S_n} \{\pi(s, \mathsf{st}(y), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \mathsf{st}(y)) \tag{A.57}$$

$$= \int_S \{\pi(s, \mathsf{st}(y), s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \mathsf{st}(y)). \tag{A.58}$$

for all $y \in {}^*X$. Thus, we have $x \in \arg\max_{\hat{x}\in X} \int_S \{\pi(s, \hat{x}, s') + \delta V(s')\}\bar{Q}_{\nu_p}(\mathrm{d}s'|s, \hat{x})$, which implies that $x$ is optimal given $s$ in the MDP($\bar{Q}_{\nu_p}$). □

## B. References


Anderson, R. M. (1982). "Star-finite representations of measure spaces". *Trans. Amer. Math. Soc.* 271.2, pp. 667–687.

– (1985). "Strong core theorems with nonconvex preferences". *Econometrica* 53.6, pp. 1283–1294.

Anderson, R. M., H. Duanmu, M. A. Khan, and M. Uyanik (2021a). "Walrasian equilibrium theory with and without free-disposal: theorems and counterexamples in an infinite-agent context". *Economic Theory*, pp. 1–26.

– (2022). "On abstract economies with an arbitrary set of players and action sets in locally-convex topological vector spaces". *Journal of Mathematical Economics* 98.

Anderson, R. M., H. Duanmu, and A. Smith (2018). "Mixing times and hitting times for general Markov processes". *Israel Journal of Mathematics, under revision.*

– (2021b). "Mixing and average mixing times for general Markov processes". *Canad. Math. Bull.* 64.3, pp. 541–552.

– (2021c). "Mixing and hitting times for Gibbs samplers and other non-Feller processes". *Illinois J. Math.* 65.3, pp. 547–577.

Anderson, R. M. and R. C. Raimondo (2008). "Equilibrium in continuous-time financial markets: Endogenously dynamically complete markets". *Econometrica* 76.4, pp. 841–907.

Anderson, R. M. and S. Rashid (1978). "A nonstandard characterization of weak convergence". *Proc. Amer. Math. Soc.* 69.2, pp. 327–332.

Arkeryd, L. O., N. J. Cutland, and C. W. Henson, eds. (1997). *Nonstandard analysis: Theory and applications in Proceedings of the NATO Advanced Study Institute on Nonstandard Analysis and its Applications.* Vol. 493. Kluwer Academic Publishers Group, Dordrecht.

Arrow, K. J. and J. R. Green (1973). "Notes on expectations equilibria in Bayesian settings". *Institute for Mathematical Studies in the Social Sciences, Unpublished.*

Berger, L. and M. Marinacci (2020). "Model uncertainty in climate change economics: A review and proposed framework for future research". *Environmental and Resource Economics*, pp. 1–27.

Bhattacharya, R. and M. Majumdar (2007). *Random dynamical systems: theory and applications.* Cambridge University Press.

Brown, D. J. and A. Robinson (1975). "Nonstandard exchange economies". *Econometrica* 43, pp. 41–55.

Cerreia-Vioglio, S., L. P. Hansen, F. Maccheroni, and M. Marinacci (2022). "Making decisions under model misspecification". *University of Chicago, Becker Friedman Institute for Economics Working Paper* 2020-103.

Cerreia-Vioglio, S., F. Maccheroni, and D. Schmeidler (2020). "Equilibria of nonatomic anonymous games". *arXiv:2005.01839.*

Cutland, N. J., V. Neves, F. Oliveira, and J. Sousa-Pinto, eds. (1995). *Developments in nonstandard mathematics.* Vol. 336. Longman, Harlow.

Duanmu, H. (2018). *Applications of Nonstandard Analysis to Markov Processes and Statistical Decision Theory.* Thesis (Ph.D.)–University of Toronto (Canada).

Duanmu, H., J. Rosenthal, and W. Weiss (2021a). "Ergodicity of Markov processes via nonstandard analysis". *Mem. Amer. Math. Soc.* 273.1342, pp. v+114.

Duanmu, H. and D. M. Roy (2021). "On extended admissible procedures and their nonstandard Bayes risk". *The Annals of Statistics* 49.4, pp. 2053–2078.

Duanmu, H., D. M. Roy, and A. Smith (2021b). "Existence of matching priors on compact spaces yielding confidence intervals". Biometrika, under revision.

Duffie, D., L. Qiao, and Y. Sun (2018). "Dynamic directed random matching". *J. Econom. Theory* 174, pp. 124–183.

Duffie, D. and Y. Sun (2007). "Existence of independent random matching". *Ann. Appl. Probab.* 17.1, pp. 386–419.

Esponda, I. and D. Pouzo (2021). "Equilibrium in misspecified Markov decision processes". *Theoretical Economics* 16, pp. 717–757.

Farmer, L., E. Nakamura, and J. Steinsson (2021). *Learning about the long run.* Tech. rep. National Bureau of Economic Research.

Frick, M., R. Iijima, and Y. Ishii (2020). "Misinterpreting others and the fragility of social learning". *Econometrica* 88.6, pp. 2281–2328.

Fudenberg, D., G. Lanzani, and P. Strack (2021). "Limit Points of Endogenous Misspecified Learning". *Econometrica* 89.3, pp. 1065–1098.

Hansen, L. P. and T. J. Sargent (2011). *Robustness*. Princeton university press.

Karlin, S. (1962). "Stochastic models and optimal policy for selling an asset". *Studies in applied probability and management science*.

Khan, M. A. (1976). "Oligopoly in markets with a continuum of traders: an asymptotic interpretation". *J. Econom. Theory* 12.2, pp. 273–297.

Khan, M. A. and Y. Sun (2001). "Asymptotic arbitrage and asset pricing models on general index sets and on the Lebesgue continuum". *Journal of Economic Theory* 101, pp. 225–251.

Koulovatianos, C., L. J. Mirman, and M. Santugini (2009). "Optimal growth and uncertainty: Learning". *Journal of Economic Theory* 144.1, pp. 280–295.

Lorenzoni, G. (2009). "A theory of demand shocks". *American Economic Review* 99.5, pp. 2050–84.

Maitra, A. (1968). "Discounted dynamic programming on compact metric spaces". *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 211–216.

Molavi, P. (2019). "Macroeconomics with learning and misspecification: A general theory and applications". *Unpublished manuscript*.

Nyarko, Y. (1991). "Learning in mis-specified models and the possibility of cycles". *Journal of Economic Theory* 55.2, pp. 416–427.

Puterman, M. L. (1994). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Robinson, A. (1966). *Non-standard analysis*. North-Holland Publishing Co., Amsterdam, pp. xi+293.

Stokey, N. L. and R. Lucas (1989). *Recursive methods in economic dynamics*. Harvard University Press.

Willard, S. (2012). *General topology*. Courier Corporation.

Wolff, M. and P. A. Loeb, eds. (2000). *Nonstandard analysis for the working mathematician*. Vol. 510. Kluwer Academic Publishers, Dordrecht.

Zimmer, G. B. (2005). "A unifying Radon-Nikodým theorem through nonstandard hulls". *Illinois J. Math.* 49.3, pp. 873–883.

## C. Supplementary Material - For Online Publication

This supplementary material is divided into four sections: (i) proofs and statements that are omitted from Appendix A, (ii) asymptotic characterization of state-action frequencies, (iii) a proof of our convergence result in Theorem 4, and (iv) a detailed analysis of some examples covered in the main paper.

C.1. **Omitted Proofs.** Theorems C.1 and C.2 are invoked at several instances during the proofs of the main theorems in our paper. We list them for completeness here.

**Theorem C.1** (Arkeryd et al. (1997, Section. 4, Corollary 6.1))**.** *Suppose* $(\Omega, \mathcal{A}, P)$ *is an internal probability space, and* $F : \Omega \to {}^*\mathbb{R}$ *is an internally integrable function such that* $\mathsf{st}(F)$ *exists everywhere. Then* $F$ *is S-integrable.*

**Theorem C.2** (Arkeryd et al. (ibid., Section. 4, Theorem 6.2))**.** *Suppose* $(\Omega, \mathcal{A}, P)$ *is an internal probability space, and* $F : \Omega \to {}^*\mathbb{R}$ *is an internally integrable function such that* $\mathsf{st}(F)$ *exists* $\overline{P}$*-almost surely. Then the following are equivalent:*

*(1)* $\mathsf{st}(\int |F| \mathrm{d}P)$ *exists and it equals to* $\lim_{n\to\infty} \mathsf{st}(\int |F_n| \mathrm{d}P)$ *where for* $n \in \mathbb{N}$, $F_n = \min\{F, n\}$ *when* $F \geq 0$ *and* $F_n = \max\{F, -n\}$ *when* $F \leq 0$*;*
*(2) For every infinite* $K > 0$, $\int_{|F|>K} |F| \mathrm{d}P \approx 0$*;*
*(3)* $\mathsf{st}(\int |F| \mathrm{d}P)$ *exists, and for every* $B$ *with* $P(B) \approx 0$*, we have* $\int_B |F| \mathrm{d}P \approx 0$*;*
*(4)* $F$ *is S-integrable with respect to* $P$*.*

We next provide a proof to Claim A.16 which is used to prove Theorem A.15.

***Proof of Claim A.16.*** Pick $(s, x) \in T_S \times T_X$ such that $Q(\mathsf{st}(s), \mathsf{st}(x))$ is dominated by $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))$. Then, $D_{\mathsf{st}(\theta)}(\cdot|\mathsf{st}(s), \mathsf{st}(x))$ is the density function of $Q(\mathsf{st}(s), \mathsf{st}(x))$ with respect to $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))$. Let $f : S \to \mathbb{R}$ be $f(t) = \ln\big(D_{\mathsf{st}(\theta)}(t|\mathsf{st}(s), \mathsf{st}(x))\big)$. For $n \in \mathbb{N}$, define $f_n : S \to \mathbb{R}$ to be:

- $f_n(t) = f(t)$ if $\frac{1}{n} \leq D_{\mathsf{st}(\theta)}(t|\mathsf{st}(s), \mathsf{st}(x)) \leq n$;
- $f_n(t) = \frac{1}{n}$ if $D_{\mathsf{st}(\theta)}(t|\mathsf{st}(s), \mathsf{st}(x)) < \frac{1}{n}$;
- $f_n(t) = n$ if $D_{\mathsf{st}(\theta)}(t|\mathsf{st}(s), \mathsf{st}(x)) > n$.

Note that $f_n$ is a bounded continuous function. Moreover, by Item 5 of Definition 2.4, we have $E_{Q(\cdot|\mathsf{st}(s),\mathsf{st}(x))} = \lim_{n\to\infty} \int_S f_n(t) Q(\mathrm{d}t|\mathsf{st}(s), \mathsf{st}(x))$. Let $F : T_S \to {}^*\mathbb{R}$ be $F(t) = \ln\big(\frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)}\big)$. For $n \in \mathbb{N}$, define $F_n : T_S \to {}^*\mathbb{R}$ to be

- $F_n(t) = F(t)$ if $\frac{1}{n} \leq \frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} \leq n$;
- $F_n(t) = \frac{1}{n}$ if $\frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} < \frac{1}{n}$;
- $F_n(t) = \frac{1}{n}$ if $\frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(s|t,x)} > n$.

By Theorem A.13, we know that $F_n(t) \approx f_n(\mathsf{st}(t))$ for every $n \in \mathbb{N}$ and $t \in T_S$. For every $n \in \mathbb{N}$, we have

$$\int_S f_n(t) Q(\mathrm{d}t|\mathsf{st}(s), \mathsf{st}(x)) = \int_{^*S} {^*f_n}(t) {^*Q}(\mathrm{d}t|\mathsf{st}(s), \mathsf{st}(x)) \tag{C.1}$$

$$\approx \int_{^*S} {^*f_n}(t) {^*Q}(\mathrm{d}t|s, x) \tag{C.2}$$

$$\approx \sum_{i \in T_S} F_n(i) \mathbb{Q}(i|s, x) \tag{C.3}$$

Thus, to finish the proof, it remains to show that $\lim_{n\to\infty} \mathsf{st}\big(\sum_{t\in T_S} F_n(i)\mathbb{Q}(t|s,x)\big) \approx \mathbb{E}_{\mathbb{Q}(\cdot|s,x)}[F(t)]$. By Theorem C.2, this is the same as establishing the S-integrability of $F(t)$ under $\mathbb{Q}(\cdot|s,x)$. Pick an infinite $K > 0$ and let $I_K = \{t \in T_S : |F(t)| > K\}$. Let $I_K^0 = \{t \in T_S : |F(t)| > K \wedge \frac{\mathbb{Q}(i|s,x)}{\mathbb{Q}_\theta(i|s,x)} \leq 1\}$ and $I_K^\infty = \{t \in T_S : |F(t)| > K \wedge \frac{\mathbb{Q}(i|s,x)}{\mathbb{Q}_\theta(i|s,x)} > 1\}$. It is easy to see that both $I_K^0$ and $I_K^\infty$ are internal sets and $I_K = I_K^0 \cup I_K^\infty$. For all $t \in I_K^0$, we have $\frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} \approx 0$. Then we have

$$\sum_{t \in I_K^0} |F(t)| \mathbb{Q}(t|s,x) = \sum_{t \in I_K^0} |F(t)| \frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} \mathbb{Q}_\theta(t|s,x) \approx 0. \tag{C.4}$$

For all $t \in I_K^\infty$, $\frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} > n$ for all $n \in \mathbb{N}$. By Theorem A.13, $D_{\mathsf{st}(\theta)}(\mathsf{st}(t)|\mathsf{st}(s), \mathsf{st}(x)) = \infty$ for all $t \in I_K^\infty$. This implies that $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))\big(\mathsf{st}(I_K^\infty)\big) = 0$. By Lemma A.11, we conclude that ${^*Q_\theta}(s,x)(\bigcup_{t \in I_K^\infty} B_S(t)) \approx 0$. By Item 5 in Definition 2.4, we conclude that $\int_{\bigcup_{t\in I_K^\infty} B_S(t)} \big({^*D_\theta}(s'|s,x)\big)^{1+r} {^*Q_\theta}(s,x)(\mathrm{d}s') \approx 0$. This implies that

$$\sum_{t \in I_K^\infty} |F(t)| \mathbb{Q}(t|s,x) = \sum_{t \in I_K^\infty} |F(t)| \frac{\mathbb{Q}(t|s,x)}{\mathbb{Q}_\theta(t|s,x)} \mathbb{Q}_\theta(t|s,x) \approx 0. \tag{C.5}$$

Combining Eq. (C.4) and Eq. (C.5), we have the desired result.

□

***Proof of Lemma A.29***. Pick $\theta \in T_\Theta$ such that $\mathsf{st}(\theta) \in \Theta_{m_p}$ and $(s,x) \in \mathrm{NS}({^*S}) \times {^*X}$ such that $Q(\mathsf{st}(s), \mathsf{st}(x))$ is dominated by $Q_{\mathsf{st}(\theta)}(\mathsf{st}(s), \mathsf{st}(x))$. By Lemma A.28 and the fact that ${^*Q_\theta}(s,x)({^*S_N}) \approx 1$, we have

$${^*\mathbb{E}}_{{^*Q^N}(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right] \approx \int_{^*S_N} {^*D_\theta}(s'|s,x) \ln\left({^*D_\theta}(s'|s,x)\right) {^*Q_\theta}(s,x)(\mathrm{d}s'). \tag{C.6}$$

By Item 5 of Definition 2.4 and the fact that $s$ is near-standard, we conclude that

$$\int_{{}^*S_N} {}^*D_\theta(s'|s,x)\ln\left({}^*D_\theta(s'|s,x)\right){}^*Q_\theta(s,x)(\mathrm{d}s') \tag{C.7}$$

$$\approx \int_{{}^*S} {}^*D_\theta(s'|s,x)\ln\left({}^*D_\theta(s'|s,x)\right){}^*Q_\theta(s,x)(\mathrm{d}s'). \tag{C.8}$$

$D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right)$ is a continuous and bounded on $S_n$ for each $n\in\mathbb{N}$. So, for every $n\in\mathbb{N}$, we have

$$\int_{S_n} D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right)Q_{\mathsf{st}(\theta)}(\mathsf{st}(s),\mathsf{st}(x))(\mathrm{d}s') \tag{C.9}$$

$$\approx \int_{{}^*S_n} {}^*D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\ln\left({}^*D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right){}^*Q_\theta(s,x)(\mathrm{d}s') \tag{C.10}$$

$$\approx \int_{{}^*S_n} {}^*D_\theta(s'|s,x)\ln\left({}^*D_\theta(s'|s,x)\right){}^*Q_\theta(s,x)(\mathrm{d}s'). \tag{C.11}$$

Note that, we have

$$\lim_{n\to\infty}\int_{S_n} D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right)Q_{\mathsf{st}(\theta)}(\mathsf{st}(s),\mathsf{st}(x))(\mathrm{d}s') \tag{C.12}$$

$$= \mathbb{E}_{Q(\cdot|\mathsf{st}(s),\mathsf{st}(x))}\left[\ln\left(D_{\mathsf{st}(\theta)}(s'|\mathsf{st}(s),\mathsf{st}(x))\right)\right]. \tag{C.13}$$

By Item 5 of Definition 2.4 and the fact that $s$ is near-standard, we have

$$\lim_{n\to\infty}\mathsf{st}\Big(\int_{{}^*S_n} {}^*D_\theta(s'|s,x)\ln\left({}^*D_\theta(s'|s,x)\right){}^*Q_\theta(s,x)(\mathrm{d}s')\Big) \tag{C.14}$$

$$\approx \int_{{}^*S} {}^*D_\theta(s'|s,x)\ln\left({}^*D_\theta(s'|s,x)\right){}^*Q_\theta(s,x)(\mathrm{d}s') \tag{C.15}$$

$$\approx {}^*\mathbb{E}_{{}^*Q^N(\cdot|s,x)}\left[\ln\left(\mathbb{D}_\theta(s'|s,x)\right)\right]. \tag{C.16}$$

Hence, we have the desired result.

□

***Proof of Lemma A.40***. Let $g$ be some element in $\mathcal{L}_{B,D}[S]$. Then there exists some $E\in\mathbb{R}_{>0}$ such that $|g(s)|\leq E+(B+D)\|s\|$ for all $s\in S$. We show that $F(g)$ is a continuous function.

**Claim C.3.** *For every* $(s,x)\in\mathrm{NS}({}^*S)\times{}^*X$, ${}^*\pi(s,x,\cdot)+\delta{}^*g(\cdot)$ *is S-integrable with respect to* ${}^*\bar{Q}_{{}^*\nu_p}(s,x)(\cdot)$.

*Proof of Claim C.3.* Let $(s_0,x_0)\in\mathrm{NS}({}^*S)\times{}^*X$ be given. By Assumption 6, it is sufficient to show that $\|s'\|$ is S-integrable with respect to ${}^*\bar{Q}_{{}^*\nu_p}(s_0,x_0)(\cdot)$. By

Lemma A.39, we have:

$$\int_{^*S} \|s'\| ^*\bar{Q}_{^*\nu_p}(\mathrm{d}s'|s_0,x_0) \approx \int_{S} \|s'\| \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s_0),\mathsf{st}(x_0)) \tag{C.17}$$

$$= \lim_{n\to\infty} \mathsf{st}\Big(\int_{^*S_n} \|s'\| ^*\bar{Q}_{^*\nu_p}(\mathrm{d}s'|s_0,x_0)\Big). \tag{C.18}$$

By Theorem C.2, $\|s'\|$ is S-integrable with respect to $^*\bar{Q}_{^*\nu_p}(\mathrm{d}s'|s_0,x_0)$.

By Claim C.3 and Theorem A.9, for every $(s,x) \in \mathrm{NS}(^*S)\times {}^*X$, we have

$$\int_{^*S} \{^*\pi(s,x,s') + \delta ^*g(s')\} ^*\bar{Q}_{^*\nu_p}(\mathrm{d}s'|s,x) \tag{C.19}$$

$$\approx \lim_{n\to\infty} \mathsf{st}\Big(\int_{^*S_n} \{^*\pi(s,x,s') + \delta ^*g(s')\} ^*\bar{Q}_{^*\nu_p}(\mathrm{d}s'|s,x)\Big) \tag{C.20}$$

$$= \lim_{n\to\infty} \int_{S_n} \{\pi(\mathsf{st}(s),\mathsf{st}(x),s') + \delta g(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s),\mathsf{st}(x)) \tag{C.21}$$

$$= \int_{S} \{\pi(\mathsf{st}(s),\mathsf{st}(x),s') + \delta g(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s),\mathsf{st}(x)). \tag{C.22}$$

Hence, we have $^*F(^*g)(s) \approx F(g)(\mathsf{st}(s))$ for all $s \in \mathrm{NS}(^*S)$, so $F(g)$ is a continuous function. For every $s \in S$, by Assumption 7, we have

$$|F(g)(s)| \leq \max_{x\in X} \int_S \{A + B\max\{\|s\|,\|s'\|\} + \delta(E + (B+D)\|s'\|)\} \bar{Q}_{\nu_p}(\mathrm{d}s'|s,x) \tag{C.23}$$

$$\leq A + \delta E + \big(B + \delta(B+D)\big)C + (B+D)\|s\| \tag{C.24}$$

Hence we have the desired result. □

***Proof of Lemma A.41***. Pick $(s,x) \in \mathrm{NS}(^*S)\times {}^*X$. Note that $^*\bar{Q}_\nu(s,x)(^*S_N) \approx 1$. Thus, we have

$$\int_{^*S_N} \{^*\pi_N(s,x,s') + \delta ^*V(s')\} ^*\bar{Q}^N_\nu(\mathrm{d}s'|s,x) \approx \int_{^*S_N} \{^*\pi_N(s,x,s') + \delta ^*V(s')\} ^*\bar{Q}_\nu(\mathrm{d}s'|s,x) \tag{C.25}$$

**Claim C.4.** *$^*\pi_N(s,x,\cdot) + \delta ^*V(\cdot)$ is S-integrable with respect to $^*\bar{Q}^N_\nu(s,x)$.*

*Proof of Claim C.4.* As $^*\bar{Q}_\nu(s,x)(^*S_N) \approx 1$, by Theorem C.2, it is sufficient to show $^*\pi_N(s,x,\cdot) + \delta ^*V(\cdot)$ is S-integrable with respect to $^*\bar{Q}_\nu(s,x)(\cdot)$. By Lemma A.40, there exists $E \in \mathbb{R}_{>0}$ such that $|V(s)| \leq E + (B+D)\|s\|$ for all $s \in S$. By Assumption 6 and Assumption 7, it is sufficient to show that $\|s'\|$ is S-integrable with respect to

${}^*\bar{Q}_\nu(s,x)(\cdot)$. By Lemma A.39 and Theorem A.9, we have

$$\int_{{}^*S} \|s'\| {}^*\bar{Q}_\nu(\mathrm{d}s'|s,x) \approx \int_S \|s'\| \bar{Q}_{\nu_p}(\mathrm{d}s'|s,x) \tag{C.26}$$

$$= \lim_{n\to\infty} \mathsf{st}\Big( \int_{{}^*S_n} \|s'\| {}^*\bar{Q}_\nu(\mathrm{d}s'|s,x)\Big). \tag{C.27}$$

By Theorem C.2, $\|s'\|$ is S-integrable with respect to ${}^*\bar{Q}_\nu(s,x)(\cdot)$.

Thus, by Arkeryd et al. (1997, Section 4, Theorem 6.2) and Theorem A.9, we have:

$$\int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*V(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x) \tag{C.28}$$

$$\approx \lim_{n\to\infty} \int_{S_n} \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s), \mathsf{st}(x)) \tag{C.29}$$

$$= \int_S \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s), \mathsf{st}(x)). \tag{C.30}$$

Hence, we have the desired result.

□

***Proof of Lemma A.42***. Let $V_0$ be the restriction of ${}^*V$ to ${}^*S_N$. For all $s \in \mathrm{NS}({}^*S)$, by Lemma A.41, we have

$$\max_{x\in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta V_0(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x) \tag{C.31}$$

$$\approx \max_{x\in X} \int_S \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta V(s')\} \bar{Q}_{\nu_p}(\mathrm{d}s'|\mathsf{st}(s), x) \tag{C.32}$$

$$= V(\mathsf{st}(s)) \approx V_0(s). \tag{C.33}$$

Let $G(f)(s) = \max_{x\in {}^*X} \int_{{}^*S_N} \{{}^*\pi_N(s,x,s') + \delta {}^*f(s')\} {}^*\bar{Q}_\nu^N(\mathrm{d}s'|s,x)$ for all $f \in {}^*\mathcal{C}_0({}^*S_N)$. Note that ${}^*\mathcal{C}_0({}^*S_N)$ is a ${}^*$complete metric space under the ${}^*$metric ${}^*d_{\mathrm{unif}}$. Consider the following internal iterated process: start with $V_0$ and define a sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ by $V_{n+1} = G(V_n)$. As $\delta \in [0,1)$ and ${}^*S_N$ is a ${}^*$compact set, there exists some $K \in {}^*\mathbb{N}$ such that ${}^*d_{\sup}(V_K, V_{K+1}) < 1$. Hence the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ is a ${}^*$Cauchy sequence with respect to the ${}^*$metric ${}^*d_{\mathrm{unif}}$. As ${}^*\mathcal{C}_0({}^*S_N)$ is ${}^*$complete, the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ has a ${}^*$limit. Note that ${}^*d_{\mathrm{unif}}(G(f_1), G(f_2)) \leq {}^*d_{\mathrm{unif}}(f_1, f_2)$ for all $f_1, f_2 \in {}^*\mathcal{C}_0({}^*S_N)$. So, $G$ is a ${}^*$continuous function, hence the ${}^*$limit of the internal sequence $\{V_n\}_{n\in {}^*\mathbb{N}}$ is the ${}^*$fixed point $\mathbb{V}$. As ${}^*d_{\mathrm{unif}}(V_0, \mathbb{V}) \approx {}^*d_{\mathrm{unif}}(V_1, V_0) \approx 0$, we have ${}^*V(s) \approx \mathbb{V}(s)$ for all $s \in \mathrm{NS}({}^*S)$.

□

C.2. **Asymptotic Characterization of State-Action Frequencies.** In this subsection, we study the problem where the agent who faces a regular SMDP with compact state and action spaces, updates her belief in each period as a result of observing the current state, her action and the new state. Our aim is to show that the agent's steady state behavior is a Berk-Nash equilibrium. Throughout this section, we work with a regular SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ as in Definition 2.4. The agent who faces this regular SMDP has a prior $\mu_0 \in \Delta(\Theta)$, which is assumed to have full support. Furthermore, throughout this section, we assume that the state space $S$ is compact. We start by making the following assumption:

**Assumption 10.** There is a referencing finite measure $\lambda$ on $(S, \mathcal{B}[S])$ with full support such that

(1) For all $\theta \in \Theta$ and all $(s, x) \in S \times X$, $Q_\theta(s, x)$ is absolutely continuous with respect to $\lambda$;
(2) The density function $q_{(\theta,s,x)}(\cdot) : S \to \mathbb{R}$ of $Q_\theta(s, x)$ with respect to $\lambda$ is a jointly continuous function on $\Theta \times S \times X \times S$;
(3) For all $(\theta, s, x) \in \Theta \times S \times X$, the density function $q_{(\theta,s,x)}(s') > 0$ for all $s' \in S$.

Recall that $\Delta(\Theta)$ denote the set of probability measures on $(\Theta, \mathcal{B}[\Theta])$, endowed with the Prokhorov metric. For $(s, x, s') \in S \times X \times S$, the Bayesian operator $B(s, x, s', \cdot) : \Delta(\Theta) \to \Delta(\Theta)$ is defined as:

$$B(s, x, s', \mu)(A) = \frac{\int_A q_{(\theta,s,x)}(s')\mu(\mathrm{d}\theta)}{\int_\Theta q_{(\theta,s,x)}(s')\mu(\mathrm{d}\theta)} \tag{C.34}$$

for all $A \in \mathcal{B}[\Theta]$.

By the principle of optimality, the agent's problem can be cast recursively as:

$$W(s, \mu) = \max_{x\in X} \int_S \{\pi(s, x, s') + \delta W(s', \mu')\}\bar{Q}_\mu(\mathrm{d}s'|s, x) \tag{C.35}$$

where $\bar{Q}_\mu(s, x) = \int_\Theta Q_\theta(s, x)\mu(\mathrm{d}\theta)$ and $\mu' = B(s, x, s', \mu)$. Let $\mathcal{C}[S \times \Delta(\Theta)]$ be the set of real-valued continuous functions on $S \times \Delta(\Theta)$, equipped with the sup-norm. Assuming Assumption 10 holds, then the operator

$$L(g)(s, \mu) = \max_{x\in X} \int_S \{\pi(s, x, s') + \delta g(s', \mu')\}\bar{Q}_\mu(\mathrm{d}s'|s, x) \tag{C.36}$$

is a contraction mapping from $\mathcal{C}[S \times \Delta(\Theta)]$ to $\mathcal{C}[S \times \Delta(\Theta)]$, with the contraction factor $\delta$. Thus, by the Banach fixed point theorem, there exists a unique $W \in \mathcal{C}[S \times \Delta(\Theta)]$ that is the solution of Eq. (C.35), which we fix for the rest of this section.

**Definition C.5.** A policy function is a function $f : S \times \Delta(\Theta) \to \Delta(X)$, where $f(\cdot|s,\mu)$ is a probability measure on $X$ if she is in state $s$ and her belief is $\mu$. A policy function is optimal if, for all $s \in S$, $\mu \in \Delta(\Theta)$ and $x \in X$ such that $x$ is in the support $f(\cdot|s,\mu)$:

$$x \in \arg\max_{\hat{x}\in X} \int_S \{\pi(s,\hat{x},s') + \delta W(s', B(s,\hat{x},s',\mu))\}\bar{Q}_\mu(\mathrm{d}s'|s,\hat{x}). \tag{C.37}$$

Let $h = (s_0, x_0, \ldots, s_k, x_k, \ldots)$ be an infinite history of state-action pairs and let $\mathbb{H} = (S \times X)^{\mathbb{N}}$ be the space of infinite histories. For every $k \in \mathbb{N}$, let $\mu_k : \mathbb{H} \to \Delta(\Theta)$ denote the agent's belief at time $k$, defined recursively by $\mu_k(h) = B(s_{k-1}, x_{k-1}, s_k, \mu_{k-1}(h))$. When the context is clear, we drop $h$ from the notation.

For a fixed $h \in \mathbb{H}$, in each period $k$, there is a state $s_k$ and a belief $\mu_k$. Given a policy function $f$, the agent chooses an action randomly according to $f(\cdot|s_k,\mu_k)$. After an action $x_k$ is realized, the state $s_{k+1}$ is drawn according to the true transition probability $Q(\cdot|s_k,x_k)$. The agent then updates her belief to $\mu_{k+1}$ according to the Bayes operator. Thus, the primitives of the problem and the policy function $f$ induce a probability distribution $\mathbb{P}^f$ over $\mathbb{H}$.

For every $k \in \mathbb{N}$, we define the frequency of the state-action pairs at time $k$ to be a function $m_k : \mathbb{H} \to \Delta(S \times X)$ such that $m_k(h)(A) = \frac{1}{k}\sum_{\tau=0}^{k} \mathbf{1}_A(s_\tau, x_\tau)$ for all measurable $A \in \Delta(S \times X)$, where $\mathbf{1}_A$ denote the indicator function on $A$.

**Definition C.6.** Let $H$ be a subset of $\mathbb{H}$. The sequence $(m_k)_{k\in\mathbb{N}}$ is said to be uniformly converges to $m \in \Delta(S \times X)$ on $H$ in total variation distance if, for every $\epsilon > 0$, there exists $k_\epsilon \in \mathbb{N}$ such that $\|m_k(h) - m\|_{\mathrm{TV}} < \epsilon$ for all $h \in H$ and all $k \geq k_\epsilon$.

*Remark* C.7. For $h \in \mathbb{H}$, the frequency of state-action pairs $m_k(h)$ is supported on a countable set. So, Definition C.6 implies that the support of $m$ is also countable. In conclusion, Definition C.6 is a reasonable assumption if both the state space $S$ and the action space $X$ are countable.

We now introduce the concept of identification and then present the main result Theorem 4 of this section.

**Definition C.8.** A SMDP is identified given $m \in \Delta(S \times X)$ if $\theta, \theta' \in \Theta_Q(m)$ implies $Q_\theta(\cdot|s,x) = Q_{\theta'}(\cdot|s,x)$ for all $(s,x) \in S \times X$.

**Theorem 4.** *Suppose Assumption 10 holds and the state space $S$ is compact. Let $f$ be an optimal policy function. Suppose:*

*(1) $(m_k)_{k\in\mathbb{N}}$ uniformly converges to some $m \in \Delta(S \times X)$ on some $H \in \mathbb{H}$ with $\mathbb{P}^f$-positive probability in total variation distance;*

*(2) The SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$ is identified given $m$.*

*Then $m$ is a Berk-Nash equilibrium for the SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle,\ \mathcal{Q}_\Theta)$.*

In this section, we present the necessary nonstandard framework to prove Theorem 4. Throughout this section, We work with a regular SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ with a compact state space $S$.

**Lemma C.9.** *Suppose Assumption 10 holds. Then the Bayesian operator $B$ is a continuous function from $S \times X \times S \times \Delta(\Theta)$ to $\Delta(\Theta)$.*

Since we are working with a regular SMDP, we can construct an associate hyperfinite SMDP $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ as in Section 4.3, which will be fixed for the rest of the section. Let $\lambda$ be the finite measure on $S$ as in Assumption 10. Define ${}^*\lambda_{T_S}$ to be the internal probability measure on $T_S$ such that ${}^*\lambda_{T_S}(\{s\}) = {}^*\lambda(B_S(s))$ for all $s \in T_S$. Let ${}^*\Delta(T_\Theta)$ denote the set of internal probability measures on $T_\Theta$. For $(s, x, s') \in T_S \times T_X \times T_S$, the hyperfinite Bayesian operator $\mathbb{B}(s, x, s', \cdot) : {}^*\Delta(T_\Theta) \to {}^*\Delta(T_\Theta)$ is given by

$$\mathbb{B}(s, x, s', \mu)(A) = \frac{\sum_{\theta\in A} \mathbb{Q}_\theta(s'|s, x)\mu(\{\theta\})}{\sum_{\theta\in T_\Theta} \mathbb{Q}_\theta(s'|s, x)\mu(\{\theta\})} \tag{C.38}$$

for all internal $A \subset T_\Theta$.

By the transfer principle and the principle of optimality, the agent's problem can be cast recursively as

$$\mathbb{W}(s, \mu) = \max_{x\in T_X} \sum_{s'\in T_S} \{\Pi(s, x, s') + \delta\mathbb{W}(s', \mu')\}\bar{\mathbb{Q}}_\mu(s'|s, x) \tag{C.39}$$

where $\bar{\mathbb{Q}}_\mu(s, x) = \sum_{\theta\in T_\Theta} \mathbb{Q}_\theta(s, x)\mu(\{\theta\})$, $\mu' = \mathbb{B}(s, x, s', \mu)$ and $\mathbb{W} : T_S \times {}^*\Delta(T_\Theta) \to {}^*\mathbb{R}$ is the unique solution to the hyperfinite Bellman equation Eq. (C.39). The existence of such a $\mathbb{W}$ is guaranteed by the transfer principle. The next theorem establishes a tight connection between the solution $\mathbb{W}$ of the hyperfinite Bellman equation Eq. (C.39) and the solution $W$ of the standard Bellman equation Eq. (C.35).

**Theorem C.10.** *Suppose Assumption 10 holds. For all $(s, \mu) \in T_S \times {}^*\Delta(T_\Theta)$:*

$$\mathbb{W}(s, \mu) \approx W(\mathsf{st}(s), \mu_p). \tag{C.40}$$

We now give the definition of hyperfinite policy functions.

**Definition C.11.** A hyperfinite policy function is an internal function $f : T_S \times {}^*\Delta(T_\Theta) \to {}^*\Delta(T_X)$, where $f(x|s, \mu)$ denotes the probability that the agent chooses $x$ if she is in state $s$ and her belief is $\mu$.

We now discuss the agent's belief updating according to the hyperfinite SMDP $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$. Recall that the agent who faces the regular SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ has a prior $\mu_0 \in \Delta(\Theta)$, which is assumed to have full support. Let $\nu_0(\theta) = {}^*\mu_0(B_\Theta(\theta))$ for all $\theta \in T_\Theta$. As $\mu_0$ has full support, then $\nu_0(\theta) > 0$ for all $\theta \in T_\Theta$. The agent who faces the hyperfinite SMDP $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ has the prior $\nu_0$. Let $h = (s_0, x_0, \ldots, s_k, x_k, \ldots)$ be an ${}^*$infinite hyperfinite history of state-action pairs and let ${}^*\mathbb{H}_{T_S \times T_X} = (T_S \times T_X)^{{}^*\mathbb{N}}$ be the space of infinite histories. It is clear that ${}^*\mathbb{H}_{T_S \times T_X} \subset {}^*\mathbb{H} = ({}^*S \times {}^*X)^{{}^*\mathbb{N}}$. For two $h_1, h_2 \in {}^*\mathbb{H}$, we write $h_1 \approx h_2$ if every coordinates of $h_1$ and $h_2$ are infinitely close. For every $k \in {}^*\mathbb{N}$, let $\nu_k : {}^*\mathbb{H}_{T_S \times T_X} \to {}^*\Delta(T_\Theta)$ denote the agent's hyperfinite belief at time $k$, defined recursively by $\nu_k(h) = \mathbb{B}(s_{k-1}, s_{k-1}, s_k, \mu_{k-1}(h))$. When the context is clear, we drop $h$ from the notation. Recall that we use $d_P$ to denote the Prokhorov metric on $\Delta(\Theta)$.

For a fixed $h \in {}^*\mathbb{H}_{T_S \times T_X}$, in each period $k$, there is a state $s_k$ and a belief $\nu_k$. Given a hyperfinite policy function $F$, the agent chooses an action randomly according to $F(\cdot|s_k, \nu_k)$. After an action $x_k$ is realized, the state $s_{k+1}$ is drawn according to the true hyperfinite transition probability $\mathbb{Q}(\cdot|s_k, x_k)$. The agent then updates her hyperfinite belief to $\nu_{k+1}$ according to the hyperfinite Bayes operator $\mathbb{B}$. Thus, the primitives of the problem and the hyperfinite policy function $F$ induce an internal probability measure ${}^*\mathbb{P}^F_{T_S \times T_X}$ over ${}^*\mathbb{H}_{T_S \times T_X}$.

For every $k \in {}^*\mathbb{N}$, we define the hyperfinite frequency of the state-action pairs at time $k$ to be a function $M_k : {}^*\mathbb{H}_{T_S \times T_X} \to {}^*\Delta(T_S \times T_X)$ such that

$$M_k(h)(\{(s,x)\}) = \frac{1}{k} \sum_{\tau=0}^{k} \mathbf{1}_{(s,x)}(s_\tau, x_\tau), \tag{C.41}$$

where $\mathbf{1}_{(s,x)}$ denote the indicator function on the point $(s, x)$.

Recall that $\mathbb{K}_\mathbb{Q} : {}^*\Delta(T_S \times T_X) \times T_\Theta \to {}^*\mathbb{R}_{\geq 0}$ denote the hyperfinite weighted Kullback Leibler divergence, and the set of closest parameter values given $m \in {}^*\Delta(T_S \times T_X)$ is the set

$$T^\mathbb{Q}_\Theta(m) = \operatorname*{arg\,min}_{\theta \in T_\Theta} \mathbb{K}_\mathbb{Q}(m, \theta). \tag{C.42}$$

The set of almost closest parameter values given $m \in {}^*\Delta(T_S \times T_X)$ is the external set

$$\hat{T}^\mathbb{Q}_\Theta(m) = \{\hat{\theta} \in T_\Theta : \mathbb{K}_\mathbb{Q}(m, \hat{\theta}) \approx \min_{\theta \in T_\Theta} \mathbb{K}_\mathbb{Q}(m, \theta)\}. \tag{C.43}$$

We now introduce the concept of S-identification for hyperfinite SMDP.

**Definition C.12.** The hyperfinite SMDP $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta\rangle, \mathscr{Q}_{T_\Theta})$ is S-identified given $m \in {}^*\Delta(T_S \times T_X)$ if $\theta, \theta' \in \hat{T}^{\mathbb{Q}}_\Theta(m)$ implies that $\mathbb{Q}_\theta(\cdot|s,x) \approx \mathbb{Q}_{\theta'}(\cdot|s,x)$ for all $(s,x) \in T_S \times T_X$.

The rigorous proof of Theorem 4 is provided in the following section.

C.3. **Proof of Theorem 4.** In this appendix, we present a rigorous proof of Theorem 4 via the hyperfinite SMDP constructed in the last section. We start by proving the continuity of Bayesian operator.

***Proof of Lemma C.9***. Note that $S \times X \times S \times \Delta(\Theta)$ is a compact metric space. Pick $(s, x, s', \mu) \in {}^*S \times {}^*X \times {}^*S \times {}^*\Delta({}^*\Theta)$. Then, $\mu$ is an internal probability measure on ${}^*\Theta$. The standard part of $\mu$ in $\Delta(\Theta)$ with respect to the Prokhorov metric is simply the push-down of $\mu$, denoted by $\mu_p$. By Assumption 10 and Theorem A.9, we have

$$\int_\Theta q_{(\theta,\mathsf{st}(s),\mathsf{st}(x))}(\mathsf{st}(s'))\mu_p(\mathrm{d}\theta) \approx \int_{{}^*\Theta} {}^*q_{(\theta,s,x)}(s')\mu(\mathrm{d}\theta). \tag{C.44}$$

Pick a set $A \in \mathcal{B}[S]$ such that $A$ is a continuity set of $B(\mathsf{st}(s), \mathsf{st}(x), \mathsf{st}(s'), \mu_p)$. Then, by Assumption 10, $A$ is a continuity set of $\mu_p$, which implies that $\mu_p(A) \approx \mu({}^*A)$. Hence, by Assumption 10 and Theorem A.9 again, we have

$$\int_A q_{(\theta,\mathsf{st}(s),\mathsf{st}(x))}(\mathsf{st}(s'))\mu_p(\mathrm{d}\theta) \approx \int_{{}^*A} {}^*q_{(\theta,s,x)}(s')\mu(\mathrm{d}\theta), \tag{C.45}$$

completing the proof. □

C.3.1. ***The Proof and Consequences of Theorem C.10***. The proof of Theorem C.10 relies on the following two lemmas:

**Lemma C.13.** *Suppose Assumption 10 holds. For all $\theta \in T_\Theta$ and all $(s, x, s') \in T_S \times T_X \times T_S$:*

$$q_{(\mathsf{st}(\theta),\mathsf{st}(s),\mathsf{st}(x))}(\mathsf{st}(s')) \approx \frac{\mathbb{Q}_\theta(s'|s,x)}{{}^*\lambda_{T_S}(\{s'\})}. \tag{C.46}$$

*Proof.* Pick $\theta \in T_\Theta$ and $(s, x, s') \in T_S \times T_X \times T_S$. Note that $\lambda$ has full support. By the construction of $T_S$, we know that ${}^*\lambda_{T_S}(\{s'\}) > 0$. By the transfer principle, we have

$$\mathbb{Q}_\theta(s'|s,x) = {}^*Q_\theta(s,x)(B_S(s')) = \int_{B_S(s')} {}^*q_{(\theta,s,x)}(y){}^*\lambda(\mathrm{d}y). \tag{C.47}$$

We also have

$$\mathbb{Q}_\theta(s'|s,x) = \int_{B_S(s')} \frac{\mathbb{Q}_\theta(s'|s,x)}{{}^*\lambda_{T_S}(\{s'\})}{}^*\lambda(\mathrm{d}y). \tag{C.48}$$

By Assumption 10, we have ${}^*q_{(\theta,s,x)}(y) \approx \frac{\mathbb{Q}_\theta(s'|s,x)}{{}^*\lambda_{T_S}(\{s'\})}$ for all $y \in B_S(s')$. By Assumption 10, we have $q_{(\mathsf{st}(\theta),\mathsf{st}(s),\mathsf{st}(x))}(\mathsf{st}(s')) \approx \frac{\mathbb{Q}_\theta(s'|s,x)}{{}^*\lambda_{T_S}(\{s'\})}$. □

Let $d_P$ denote the Prokhorov metric on $\Delta(\Theta)$. By the transfer principle, ${}^*d_P$ is the *Prokhorov metric on ${}^*\Delta({}^*\Theta)$.

**Lemma C.14.** *Suppose Assumption 10 holds. For all $\mu \in {}^*\Delta(T_\Theta)$ and all $(s, x, s') \in T_S \times T_X \times T_S$:*

$$ {}^*d_P\big({}^*B(\mathsf{st}(s), \mathsf{st}(x), \mathsf{st}(s'), \mu_p), \mathbb{B}(s, x, s', \mu)\big) \approx 0. \tag{C.49} $$

*That is, the hyperfinite Bayesian operator $\mathbb{B}(s, x, s', \mu)$ is in the monad of the standard Bayesian operator $B(\mathsf{st}(s), \mathsf{st}(x), \mathsf{st}(s'), \mu_p)$, with respect to the Prokhorov metric $d_P$.*

*Proof.* For all $\mu \in {}^*\Delta(T_\Theta)$, as $\Theta$ is compact, $\mu_p$ is a well-defined probability measure on $\Theta$, and $\mu$ is in the monad of $\mu_p$ with respect to the Prokhorov metric. Then the result follows from Lemma C.13 and Theorem A.9. □

We now give a rigorous proof of Theorem C.10

***Proof of Theorem C.10***. Let $W_0$ be the restriction of ${}^*W$ on $T_S \times {}^*\Delta(T_\Theta)$. For all $(\mu, s, x) \in {}^*\Delta(T_\Theta) \times T_S \times T_X$, by Lemma A.18, Lemma C.9 and Theorem A.9, we have

$$ \int_{T_S} \{\Pi(s, x, s') + \delta W_0(s', \mu')\} \bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s, x) \tag{C.50} $$

$$ \approx \int_S \{\pi(\mathsf{st}(s), \mathsf{st}(x), s') + \delta W(s', \mu'_p)\} \bar{Q}_{\mu_p}(\mathrm{d}s'|\mathsf{st}(s), \mathsf{st}(x)). \tag{C.51} $$

Thus, we can conclude that

$$ \max_{x \in T_X} \int_{T_S} \{\Pi(s, x, s') + \delta W_0(s', \mu')\} \bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s, x) \tag{C.52} $$

$$ \approx \max_{x \in X} \int_S \{\pi(\mathsf{st}(s), x, s') + \delta W(s', \mu'_p)\} \bar{Q}_{\mu_p}(\mathrm{d}s'|\mathsf{st}(s), x) \tag{C.53} $$

$$ = W(\mathsf{st}(s), \mu_p) \approx W_0(s, \mu). \tag{C.54} $$

Let

$$ \mathbb{L}(g)(s) = \max_{x \in T_X} \int_{T_S} \{\Pi(s, x, s') + \delta g(s', \mu')\} \bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s, x) \tag{C.55} $$

for all internal function $g : T_S \times {}^*\Delta(T_\Theta) \to {}^*\mathbb{R}$. Note that $\mathbb{L}$ is a contraction with the contraction factor $\delta$. Moreover, we can find $\mathbb{W}$ as following: start with $W_0$ and define a sequence $\{W_n\}_{n \in {}^*\mathbb{N}}$ by $W_{n+1} = \mathbb{H}(W_n)$. Then $\mathbb{W}$ is the *limit of $\{W_n\}_{n \in {}^*\mathbb{N}}$. Thus, we have

$$^*d_{\sup}(W_0, \mathbb{W}) \leq \frac{1}{1-\delta} {^*d_{\sup}}(W_1, W_0) \approx 0. \tag{C.56}$$

As $W$ is continuous, we have $\mathbb{W}(s,\mu) \approx W(\mathsf{st}(s), \mu_p)$ for all $(s,\mu) \in T_S \times {^*\Delta}(T_\Theta)$. □

We now prove two important consequences of Theorem C.10, which will be used in the proof of Theorem 4. Let $Y$ be an arbitrary metric space and $U$ be a subset of $^*Y$. The *nonstandard hull* of $U$, denoted by $\hat{U}$, is the collection of all points in $^*Y$ that are infinitely close to some point in $U$. That is:

$$\hat{U} = \{y \in {^*Y} : (\exists u \in U)({^*d_Y}(y,u) \approx 0)\}. \tag{C.57}$$

**Lemma C.15.** *Suppose Assumption 10 holds. Let* $(s,\mu) \in T_S \times {^*\Delta}(T_\Theta)$. *Suppose*

$$y \in \underset{x \in T_X}{\arg\hat{\max}} \int_{T_S} \{\Pi(s,x,s') + \delta\mathbb{W}(s', \mathbb{B}(s,x,s',\mu))\}\bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s,x). \tag{C.58}$$

*Then*

$$\int_{T_S} \{\Pi(s,y,s') + \delta\mathbb{W}(s', \mathbb{B}(s,y,s',\mu))\}\bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s,y) \tag{C.59}$$

$$\approx \max_{x \in T_X} \int_{T_S} \{\Pi(s,x,s') + \delta\mathbb{W}(s', \mathbb{B}(s,x,s',\mu))\}\bar{\mathbb{Q}}_\mu(\mathrm{d}s'|t,x). \tag{C.60}$$

*Proof.* Pick

$$x_0 \in \underset{x \in T_X}{\arg\max} \int_{T_S} \{\Pi(s,x,s') + \delta\mathbb{W}(s', \mathbb{B}(s,x,s',\mu))\}\bar{\mathbb{Q}}_\mu(\mathrm{d}s'|s,x) \tag{C.61}$$

such that $y \approx x_0$. As the SMDP $(\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ is regular, we have $\Pi(s,y,s') \approx \Pi(s,x_0,s')$ for all $s' \in T_S$. By regularity again, the $^*$Prokhorov distance between $\bar{\mathbb{Q}}(\cdot|s,y)$ and $\bar{\mathbb{Q}}(\cdot|s,x_0)$ is infinitesimal. The result then follows from Lemma C.14 and Theorem C.10. □

**Lemma C.16.** *Suppose Assumption 10 holds. Let* $(s_1,\mu) \in {^*S} \times {^*\Delta}({^*\Theta})$ *and* $(s_2,\nu) \in T_S \times {^*\Delta}(T_\Theta)$ *such that* $s_1 \approx s_2$ *and the* $^*$*Prokhorov distance between* $\mu$ *and* $\nu$ *is infinitesimal. Suppose*

$$x \in \underset{\hat{x} \in {^*X}}{\arg\max} \int_{^*S} \{{^*\pi}(s_1,\hat{x},s') + \delta {^*W}(s', {^*B}(s_1,\hat{x},s',\mu))\}{^*\bar{Q}_\mu}(\mathrm{d}s'|s_1,\hat{x}). \tag{C.62}$$

*Then, for all $y \in T_X$ such that $y \approx x$:*

$$\int_{T_S} \{\Pi(s_2, y, s') + \delta \mathbb{W}(s', \mathbb{B}(s_2, y, s', \nu))\} \bar{\mathbb{Q}}_\nu(\mathrm{d}s'|s_2, y) \tag{C.63}$$

$$\approx \max_{\hat{x} \in T_X} \int_{T_S} \{\Pi(s_2, \hat{x}, s') + \delta \mathbb{W}(s', \mathbb{B}(s_2, \hat{x}, s', \nu))\} \bar{\mathbb{Q}}_\nu(\mathrm{d}s'|s_2, \hat{x}). \tag{C.64}$$

*Proof.* By Lemma C.14, Theorem C.10 and the fact that $(\langle S, X, q_0, Q, \pi, \delta\rangle, \mathcal{Q}_\Theta)$ is a regular SMDP, we have

$$\arg\max_{\hat{x} \in {}^*X} \int_{{}^*S} \{{}^*\pi(s_1, \hat{x}, s') + \delta {}^*W(s', {}^*B(s_1, \hat{x}, s', \mu))\} {}^*\bar{Q}_\mu(\mathrm{d}s'|s_1, \hat{x}) \tag{C.65}$$

$$\approx \arg\max_{\hat{x} \in T_X} \int_{T_S} \{\Pi(s_2, \hat{x}, s') + \delta \mathbb{W}(s', \mathbb{B}(s_2, \hat{x}, s', \nu))\} \bar{\mathbb{Q}}_\nu(\mathrm{d}s'|s_2, \hat{x}). \tag{C.66}$$

Moreover, we have

$$y \in \arg\hat{\max}_{z \in {}^*X} \int_{{}^*S} \{{}^*\pi(s_1, z, s') + \delta {}^*W(s', {}^*B(s_1, z, s', \mu))\} {}^*\bar{Q}_\mu(\mathrm{d}s'|s_1, z) \tag{C.67}$$

$$= \arg\hat{\max}_{z \in T_X} \int_{T_S} \{\Pi(s_2, z, s') + \delta \mathbb{W}(s', \mathbb{B}(s_2, z, s', \nu))\} \bar{\mathbb{Q}}_\nu(\mathrm{d}s'|s_2, z). \tag{C.68}$$

By Lemma C.15, we have the desired result. □

C.3.2. ***Proof of Theorem 4***. We are now at the place to prove Theorem 4. We start with the following lemma, which shows that the agent's belief $\mu_k$ and the agent's hyperfinite belief $\nu_k$ remains close for some infinite steps.

**Lemma C.17.** *Suppose Assumption 10 holds. Let $\tilde{h} \in {}^*\mathbb{H}_{T_S \times T_X}$ and $h \in {}^*\mathbb{H}$ be such that $\tilde{h} \approx h$. Then, for every $k \in \mathbb{N}$, ${}^*d_P({}^*\mu_k, \nu_k) \approx 0$. Hence, there exists some $k_0 \in {}^*\mathbb{N} \setminus \mathbb{N}$ such that ${}^*d_P({}^*\mu_k, \nu_k) \approx 0$ for all $k \leq k_0$.*

*Proof.* The second claim follows from the first claim and saturation. We now prove the first claim by induction. Clearly, we have ${}^*d_P({}^*\mu_0, \nu_0) \approx 0$. The inductive case follows from Lemma C.9 and Lemma C.14. □

If the frequency of state-action pairs $(m_k)_{k \in \mathbb{N}}$ uniformly converges in total variation distance to some $m \in \Delta(S \times X)$ for all $h$ in some set $H \subset \mathbb{H}$, then the hyperfinite frequence of state-action pairs $(M_k)_{k \in {}^*\mathbb{N}}$ almost converges to some $M \in {}^*\Delta(T_S \times T_X)$ for all $\tilde{h}$ in some internal $\tilde{H} \subset {}^*\mathbb{H}_{T_S \times T_X}$. As one would expect, $M$ and $\tilde{H}$ are closely related to $m$ and $H$, respectively.

**Lemma C.18.** *Let $H \in \mathbb{H}$ be such that $(m_k(h))_{k \in \mathbb{N}}$ converges in total variation distance to some $m \in \Delta(S \times X)$ for all $h \in H$. Let $M \in {}^*\Delta(T_S \times T_X)$ be $M(\{(s, x)\}) =$*

${}^*m(B_S(s) \times B_X(x))$. *Let* $\tilde{H}$ *be the internal subset of* ${}^*\mathbb{H}_{T_S \times T_X}$ *consisting of* $\tilde{h} = (\tilde{s}_0, \tilde{x}_0, \ldots, \tilde{s}_k, \tilde{x}_k, \ldots) \in \tilde{H}$ *such that*

$$(\exists h = (s_0, x_0, \ldots, s_k, x_k, \ldots) \in {}^*H)(\forall k \in {}^*\mathbb{N})(s_k \in B_S(\tilde{s}_k) \wedge x_k \in B_X(\tilde{x}_k)). \quad \text{(C.69)}$$

*Then* $(M_k(\tilde{h}))_{k \in {}^*\mathbb{N}}$ ${}^*$*converges to* $M$ *for all* $\tilde{h} \in \tilde{H}$. *Moreover, if* $(m_k)_{k \in \mathbb{N}}$ *uniformly converges to* $m$ *on* $H$ *in total variation distance, then*

$$\|M_k(\tilde{h}) - M\|_{\mathrm{TV}} \approx 0 \quad \text{(C.70)}$$

*for all* $\tilde{h} \in \tilde{H}$ *and all* $k \in {}^*\mathbb{N} \setminus \mathbb{N}$.

*Proof.* Pick some $\tilde{h} = (\tilde{s}_0, \tilde{x}_0, \ldots, \tilde{s}_k, \tilde{x}_k, \ldots) \in \tilde{H}$. By the construction of $\tilde{H}$, there exists some $h = (s_0, x_0, \ldots, s_k, x_k, \ldots) \in {}^*H$ such that $s_k \in B_S(\tilde{s}_k)$ and $x_k \in B_X(\tilde{x}_k)$ for all $k \in {}^*\mathbb{N}$. By the transfer principle,

$$\big({}^*m_k(h)(B_S(s) \times B_X(x))\big)_{k \in {}^*\mathbb{N}} \ {}^*\text{converges to } {}^*m(B_S(s) \times B_X(x)) \quad \text{(C.71)}$$

for all $(s, x) \in T_S \times T_X$. For all $(s, x) \in T_S \times T_X$, note that $(\tilde{s}_k, \tilde{x}_k) = (s, x)$ if and only if $(s_k, x_k) \in B_S(s) \times B_X(x)$. Hence, we conclude that $\big(M_k(\tilde{h})(\{(s, x)\})\big)_{k \in {}^*\mathbb{N}}$ ${}^*$converges to $M(\{(s, x)\})$ for all $(s, x) \in T_S \times T_X$. Now, suppose that $(m_k)_{k \in \mathbb{N}}$ uniformly converges to $m$ on $H$ in total variation distance, by saturation, we have $\|{}^*m_k(h) - {}^*m\|_{\mathrm{TV}} \approx 0$ for all $h \in {}^*H$ and all $k \in {}^*\mathbb{N} \setminus \mathbb{N}$. By the construction of $\tilde{H}$, $T_S$ and $T_X$, we have the desired result. □

Using essentially the same proof as in Lemma 2 of EP, we have:

**Lemma C.19.** *Let* $F$ *be a hyperfinite policy function. Suppose that* $\|M_k(\tilde{h}) - M\|_{\mathrm{TV}} \approx 0$ *for some* $k \in {}^*\mathbb{N}$ *and all* $\tilde{h}$ *in some internal* $\tilde{H} \subset {}^*\mathbb{H}_{T_S \times T_X}$ *such that* ${}^*\mathbb{P}^F_{T_S \times T_X}(\tilde{H}) > 0$. *Then, for any internal set* $A \supset \hat{T}^{\mathbb{Q}}_{\Theta}(m)$, $\nu_k(A) \approx 1$ *on some internal set* $\tilde{H}' \subset \tilde{H}$ *with* ${}^*\mathbb{P}^F_{T_S \times T_X}(\tilde{H}') > 0$.

We now study the connection between identification and S-identification. The following lemma follows from essentially the same proof of Theorem A.17.

**Lemma C.20.** *Let* $m \in \Delta(S \times X)$ *and let* $M \in {}^*\Delta(T_S \times T_X)$ *be the same as in Lemma C.18. Then, for every* $\hat{\theta} \in \hat{T}^{\mathbb{Q}}_{\Theta}(M)$, $\mathsf{st}(\hat{\theta}) \in \Theta_Q(m)$.

The hyperfinite SMDP is S-identified if the SMDP is identified.

**Lemma C.21.** *Suppose Assumption 10 holds and the SMDP* $(\langle S, X, q_0, Q, \pi, \delta \rangle, \mathscr{Q}_{\Theta})$ *is identified given* $m \in \Delta(S \times X)$. *Let* $M \in {}^*\Delta(T_S \times T_X)$ *be the same as in Lemma C.18. Then the hyperfinite SMDP* $(\langle T_S, T_X, h_0, \mathbb{Q}, \Pi, \delta \rangle, \mathscr{Q}_{T_{\Theta}})$ *is S-identified given* $M$.

*Proof.* By Lemma C.20, for every $\theta \in \hat{T}_{\Theta}^{\mathbb{Q}}(m)$, we have $\mathsf{st}(\theta) \in \Theta_Q(m)$. Thus, for $\theta, \theta' \in \hat{T}_{\Theta}^{\mathbb{Q}}(m)$, by Assumption 10, we have

$$ {}^*Q_{\theta}(\cdot|s,x) \approx {}^*Q_{\mathsf{st}(\theta)}(\cdot|s,x) = {}^*Q_{\mathsf{st}(\theta')}(\cdot|s,x) \approx {}^*Q_{\theta'}(\cdot|s,x) \tag{C.72} $$

for all $(s,x) \in {}^*S \times {}^*X$. This immediately implies that $\mathbb{Q}_{\theta}(\cdot|s,x) \approx \mathbb{Q}_{\theta'}(\cdot|s,x)$ for all $(s,x) \in T_S \times T_X$, completing the proof. □

We now prove the main result, Theorem 4.

***Proof of Theorem 4.*** Let $M \in {}^*\Delta(T_S \times T_X)$ be the same as in Lemma C.18. Let $\tilde{H} \subset {}^*\mathbb{H}_{T_S \times T_X}$ be the same internal set as in Lemma C.18. By Lemma C.18, we have:

$$ \|M_k(\tilde{h}) - M\|_{\mathrm{TV}} \approx 0 \tag{C.73} $$

for all $\tilde{h} \in \tilde{H}$ and all $k \in {}^*\mathbb{N} \setminus \mathbb{N}$. Let $F$ be a hyperfinite policy function such that ${}^*\mathbb{P}^F_{T_S \times T_X}(\tilde{H}) > 0$. Pick $\tilde{H}' \subset \tilde{H}$ as in Lemma C.19. For the rest of the proof, we fix $\tilde{h} = (\tilde{s}_0, \tilde{x}_0, \dots, \tilde{s}_k, \tilde{x}_k, \dots) \in \tilde{H}'$. By the construction of $\tilde{H}$, there exists $h = (s_0, x_0, \dots, s_k, x_k, \dots) \in {}^*H$ such that $s_k \in B_S(\tilde{s}_k)$ and $x_k \in B_X(\tilde{x}_k)$ for all $k \in {}^*\mathbb{N}$. For every $k \in {}^*\mathbb{N}$, let ${}^*\mu_k$ denote the updated ${}^*$belief and $\nu_k$ denote the updated hyperfinite belief at time $k$, according to $h$ and $\tilde{h}$, respectively. Henceforth, we omit the hyperfinite history from the notation.

Recall that we use $d_P$ to denote the Prokhorov metric on $\Delta(\Theta)$. By Lemma C.17, there exists some $k_0 \in {}^*\mathbb{N} \setminus \mathbb{N}$ such that ${}^*d_P({}^*\mu_k, \nu_k) \approx 0$ for all $k \leq k_0$. Let $(s,x) \in T_S \times T_X$ be such that $M(\{(s,x)\}) > 0$. By the construction of $M$, $(\mathsf{st}(s), \mathsf{st}(x))$ is in the support of $m$. Thus, there exists $k_1 \in {}^*\mathbb{N} \setminus \mathbb{N}$ such that

(1) $k_1 \leq k_0$;
(2) $\tilde{s}_{k_1} \approx s_{k_1} \approx s$ and $\tilde{x}_{k_1} \approx x_{k_1} \approx x$.

As $f$ is an optimal policy function, by the transfer principle, we have

$$ x_{k_1} \in \operatorname*{arg\,max}_{\hat{x} \in {}^*X} \int_{{}^*S} \{{}^*\pi(s_{k_1}, \hat{x}, s') + \delta {}^*W(s', {}^*B(s_{k_1}, \hat{x}, s', {}^*\mu_{k_1}))\} {}^*\bar{Q}_{{}^*\mu_{k_1}}(\mathrm{d}s'|s_{k_1}, \hat{x}). \tag{C.74} $$

As ${}^*d_P({}^*\mu_{k_1}, \nu_{k_1}) \approx 0$, by Lemma C.16, we have:

$$ \int_{T_S} \{\Pi(s,x,s') + \delta \mathbb{W}(s', \mathbb{B}(s,x,s',\nu_{k_1}))\} \bar{\mathbb{Q}}_{\nu_{k_1}}(\mathrm{d}s'|s,x) \tag{C.75} $$

$$ \approx \max_{\hat{x} \in T_X} \int_{T_S} \{\Pi(s,\hat{x},s') + \delta \mathbb{W}(s', \mathbb{B}(s,\hat{x},s',\nu_{k_1}))\} \bar{\mathbb{Q}}_{\nu_{k_1}}(\mathrm{d}s'|s,\hat{x}). \tag{C.76} $$

As the SMDP is identified given $m$, by Lemma C.21, the hyperfinite SMDP is S-identified. This implies that there is $\mathbb{Q}_M$ such that, for all $\nu \in {}^*\Delta(T_\Theta)$ with support being a subset of $\hat{T}_\Theta^{\mathbb{Q}}(M)$, $\mathbb{Q}_\nu \approx \mathbb{Q}_M$. By Lemma C.19, the support of $\nu_{k_1}$ is a subset of $\hat{T}_\Theta^{\mathbb{Q}}(M)$. Note that the support of the posterior of $\nu_{k_1}$ generated from the hyperfinite Bayesian operator is a subset of the support of $\nu_{k_1}$. Hence, we have:

$$\int_{T_S} \{\Pi(s,x,s') + \delta\mathbb{V}(s')\}\mathbb{Q}_M(\mathrm{d}s'|s,x) \tag{C.77}$$

$$\approx \max_{\hat{x}\in T_X} \int_{T_S} \{\Pi(s,\hat{x},s') + \delta\mathbb{V}(s')\}\mathbb{Q}_M(\mathrm{d}s'|s,\hat{x}). \tag{C.78}$$

By Lemma C.18, $(M_k)_{k\in {}^*\mathbb{N}}$ *converges to $M$ for all hyperfinite histories in $\tilde{H}$. By the transfer principle (or use essentially the same proof as in Theorem 2 of EP, $M$ is *stationary. Thus, $M$ is a Berk-Nash S-equilibrium for the hyperfinite SMDP with the hyperfinite belief $\nu_{k_1}$ (or any $\nu \in {}^*\Delta(T_\Theta)$ such that the support of $\nu$ is a subset of $\hat{T}_\Theta^{\mathbb{Q}}(M)$). Hence, by Theorem 4.11, $m$ is a Berk-Nash equilibrium for the SMDP with the belief being any $\mu \in \Delta(\Theta)$ such that the support of $\mu$ is a subset of $\Theta_Q(m)$. □

C.4. **Detailed Analysis of Examples.** Here we present the analysis of the examples 3.1, 3.3, 3.4 and 3.5.

**Example 3.1** In this example, we study an AR(1) process and show that a Berk-Nash equilibrium exists if and only if the AR(1) process has no unit root. Recall that the SMDP in this problem is defined as:

- The state space $S = \mathbb{R}$, the action space $X = \{0\}$, and the payoff function $\pi : S \times X \times S \to \mathbb{R}$ is the constant function 0;
- For every $s \in S$, the true transition probability function $Q(s)$ is the distribution of $a_0 s + b_0\xi$, where $a_0 \in [0,2], b_0 \in [0,1]$ and $\xi = \mathcal{N}(0,1)$ has the standard normal distribution;
- he parameter space $\Theta$ is $[0,2] \times [0,1]$ and for every $(a,b) \in \Theta$, the transition probability function $Q_{(a,b)}(s)$ is the distribution of $as + b\xi$.

We first consider the degenerate case $b_0 = 0$. The true transition $Q_{(a_0,0)}$ is absolutely continuous with respect to $Q_{(a,b)}$ if and only if $a = a_0$ and $b = b_0 = 0$. When $a_0 < 1$, the Markov process has a unique stationary distribution, namely the Dirac measure $\delta_0$ at zero. So the Berk-Nash equilibrium is $\delta_{(0,0)}$ with the belief $\delta_{(a_0,0)}$. When $a_0 = 1$, the Dirac measure $\delta_s$ is a stationary distribution for every $s \in S$, and $\delta_{(s,0)}$ is a Berk-Nash equilibrium supported by the belief $\delta_{(1,0)}$. When $a_0 > 1$, there is no stationary distribution hence no Berk-Nash equilibrium.

For the non-degenerate case $b_0 > 0$, following Example 3.1, we focus on the case $0 \leq a_0 < 1$. We now provide rigorous verification for Assumption 2 and Assumption 3:

- We apply the Lyapunov condition to verify Assumption 2 by taking the Lyapunov function $V(s) = |s|$. Clearly, this $V$ is a non-negative, continuous and norm-like function as defined in Assumption 2. Moreover, we have $S_n = \{s \in S : V(s) \leq n\}$ for all $n \in \mathbb{N}$. By the properties of the folded normal distribution, we have:

$$\int_S |s'|Q(s)(\mathrm{d}s') = b_0\sqrt{\frac{2}{\pi}}e^{-\frac{a_0^2 s^2}{2b_0^2}} + a_0 s(1 - 2\phi(-\frac{a_0 s}{b_0})) \tag{C.79}$$

  where $\phi$ is the cumulative distribution function of the standard normal distribution. Thus, for all $s \geq 0$, we have $\int_S |s'|Q(s)(\mathrm{d}s') \leq a_0|s| + \sqrt{\frac{2}{\pi}}$. Hence, by choosing $\alpha = 1 - a_0$ and $\beta = \sqrt{\frac{2}{\pi}}$, Eq. (2.12) is satisfied. Hence, Assumption 2 is satisfied;

- For $(a, b) \in [0, 2] \times (0, 1]$ and $s \in S$, the relative entropy from $Q_{(a,b)}(s)$ to $Q(s)$ is:

$$\mathcal{D}_{\mathrm{KL}}(Q(s), Q_{(a,b)}(s)) = \ln(\frac{b}{b_0}) + \frac{b_0^2 + (a_0 s - as)^2}{2b^2} - \frac{1}{2}. \tag{C.80}$$

  Note that the true transition probability function $Q(s)$ has a unique stationary measure $\mu = \mathcal{N}(0, \frac{b_0^2}{1-a_0^2})$. It is then straightforward to show that Assumption 3 is satisfied.

**Example 3.3** This example assumes that the agent knows the per-period payoff function and the transition function but has a misspecified revenue function. We follow EP in framing the price shock be a part of the state variable. The Bellman can be written as,

$$V(z, \epsilon) = \max_x \int_{Z\times[0,1]} (zf(x)\epsilon' - c + \delta V(z', \epsilon'))\, Q(dz' \mid z)\, Q^R(d\epsilon' \mid x)$$

The variable $\epsilon'$ is the unknown price shock to the revenue, $r = f(x)\epsilon'$, at the time the agent has to choose $x$. Its distribution is given by $Q^R(d\epsilon' \mid x) \sim d_\theta$. The agent knows $Q$, but does not know $Q^R$. In particular, the agent has a parametric family of transitions, where $Q^R_\theta(d\epsilon' \mid x)$ is the distribution of $\epsilon'$. The parameter space $\Theta$ is compact, that is, $\Theta = [0, 2KE_{d^*}[\epsilon] + 1]$ and the action space, $X = [0, \max\left[\left((E_{d^*}[\epsilon])/4\right)^{2/3}, \left((E_{d^*[\epsilon]})/\sqrt{K}\right)^{2/3}\right] + 1]$, where $K = (1 - e^{-k})/(1 - k(k +$

$1)e^{-k}$). Given this, suppose the true production function is given by $f^*(x)x^{1/2}$ and hence, concave. Then the minimizer, $\theta^* = 2(K\mathbb{E}_{d^*}[\epsilon])^{2/3}$ and the corresponding optimal action for the misspecified agent is $x^* = z\theta^*/2K$,[35] whereas, for the agent with the correctly specified model is, $x^{opt} = \left(\frac{z\mathbb{E}_{d^*}[\epsilon]}{4}\right)^{2/3}$. We first solve for the optimal action as a function of model primitives. Suppose the agent has a degenerate belief on some $\theta$. Here, as in the original example, the agent's optimization problem reduces to a static optimization problem $\max_x zxE_\theta[\epsilon] - x^2$. Noting that $E_\theta[\epsilon] = \frac{\theta}{K}$,[36] it follows that the optimal input choice in state $z$ is $x^* = \frac{z\theta}{2K}$. Next, the stationarity condition implies that the marginal of $m$ over $\mathbb{Z}$ is equal to the stationary distribution over $z$, which is $q$, a uniform distribution, $U[0,1]$. Therefore, the stationary distribution over $\mathbb{X}$, denoted by $m_{\mathrm{X}}$, has a uniform support over $[0, \frac{\theta}{2K}]$. Finally, we optimize $\theta$ for the weighted KLD,

$$\int_x E_{Q(\cdot|x)}\left[\log Q_\theta^R(f' \mid x)\right] m_{\mathrm{X}}(x)dx = \int_{[0,1]} E_{Q(\cdot|x)}\left[\log d_\theta(\epsilon')\right] m_{\mathrm{X}}(x)dx$$
$$= \int_{[0,1]} E_{Q(\cdot|x)}\left(-\frac{1}{\theta}(\epsilon') - \ln\theta - \ln(1-\exp^{-k})\right) m_{\mathrm{X}}(.)dx$$

Then minimizing the above expression with respect to $\theta$ gives us the minimizing $\theta^*$ and the corresponding $x^*$.

**Example 3.4** For this optimal savings problem, we solve for optimality, belief restriction and stationarity. The Bellman equation for the agent is

$$V(y,z) = \max_{0\leq x\leq y} z\ln(y-x) + \delta E\left[V(y', z') \mid x\right],$$

and let us guess that the form of the value function is $V(y,z) = a(z) + b(z)\ln(y)$. This provides us a guess for the optimal strategy which is to invest a fraction of wealth that depends on the utility shock and the unknown parameter $\beta$, i.e., $x = A_z(\beta)\cdot y$, where $A_z(\beta) = \frac{\delta\beta E[b(z')]}{(z+\delta\beta E[b(z')])}$ where $b(z)$ satisfies $b(z) = z+\delta\beta E[b(z')]$ for $z \in [0,1]$. Solving for $b(z)$, we get $b(z) = z + \frac{\delta\beta}{1-\delta\beta}E[z]$ which gives $A_z(\beta) = \frac{0.5\delta\beta}{(1-\delta\beta)z + 0.5\delta\beta}$ where $E[z] = 0.5$.[37] The stationarity condition is met because of $0 \leq \beta^* < 1$, which

---

[35] For $k > 0$, let $K = \frac{1-e^{-k}}{1-k(k+1)e^{-k}}$ which is always finite and asymptotes to 1 as $k \to \infty$.

[36] $K = \frac{1-e^{-k}}{1-k(k+1)e^{-k}}$

[37] This corrects the typo in EP for the policy function.

prevents the process from drifting away. The belief restriction and the rest of the problem for $\beta^m$ is solved analogously as in EP.

**Example 3.5** This example assumes that the agent knows the per-period payoff function and the transition function but has a misspecified cost function. We follow EP in framing cost be a part of the state variable. we simply let the cost $c$ be part of the state as follows:

$$V(z,c)=\max_x \int_{Z\times C} \left(zf(x)-c'+\delta V\left(z',c'\right)\right) Q\left(dz'\mid z\right) Q^C\left(dc'\mid x\right)$$

The variable $c'$ is the unknown cost of production at the time the agent has to choose $x$. Its distribution is given by $Q^C\left(dc'\mid x\right)$, which is the distribution of $c'=c(x)$ as described above. The agent knows $Q$, but does not know $Q^C$. In particular, the agent has a parametric family of transitions, where $Q_\theta^C\left(dc'\mid x\right)$ is the distribution of $c'=c_\theta(x)$. The action space, $X=\left[\varepsilon,\max\left[\left((\mathbb{E}_{d^*}[\epsilon])/4\right)^{2/3},\left((\mathbb{E}_{d^*}[\epsilon])/\sqrt{K}\right)^{2/3}\right]+1\right], \varepsilon>0$. The parameter space $\Theta$ is compact, $\Theta=\left[0,\sqrt{\frac{K+1}{2}\mathbb{E}_{d^*}[\epsilon]+1}\right]$, where $K=\dfrac{1-e^{-k}}{1-k(k+1)e^{-k}}$. Given this, suppose the true cost function $\phi(x)$ is quadratic i.e. $\phi(x)=x^2$. Then the Berk-Nash equilibrium is characterized by the minimizer $\theta^*$ given by $\theta^*=\sqrt{\frac{K}{2}\mathbb{E}_{d^*}[\epsilon]}$ and the action $x^*=\sqrt{\frac{2K}{\mathbb{E}_{d^*}(\epsilon)}}z$, whereas for the agent with the correctly specified model is $x^{opt}=\dfrac{\sqrt{z}}{\sqrt{2\mathbb{E}_{d^*}(\epsilon)}}$. Indeed, note that the true transition probability function $Q(s)$ has a unique stationary measure $\mu$. Therefore, the Berk-Nash equilibrium for this SMDP is $\mu\times\delta_{x^*}$, supported by the belief $\delta_{\theta^*}$.

We now solve for the equilibrium. First, we solve for the optimal $x*$ as a function of the parameter. Suppose the agent has a degenerate belief on some $\theta$. Here, as in the original example, the agent's optimization problem reduces to a static optimization problem $\max_x z\ln x - xE_\theta[\epsilon]$. Noting that $E_\theta[\epsilon]=\dfrac{\theta}{K}$, it follows that the optimal input choice in state $z$ is $x^*=Kz/\theta$. Next, the stationarity condition implies that the marginal of $m$ over $\mathbb{Z}$ is equal to the stationary distribution over $z$, which is $q$, a uniform distribution, $U[0,1]$. Therefore, the stationary distribution over $\mathbb{X}$, denoted by $m_{\mathrm{X}}$, is a uniform distribution, $U[0,\frac{K}{\theta}]$. Finally, following the steps as in Example 3.3, we get our corresponding $x^*$ and $\theta^*$.

Robert M. Anderson, University of California, Berkeley, Departments of Economics and Mathematics

Haosui Duanmu, Harbin Institute of Technology, Department of Mathematics

Aniruddha Ghosh, The Johns Hopkins University, Department of Economics

M. Ali Khan, The Johns Hopkins University, Department of Economics