Duality and Estimation of Undiscounted Markov Decision Processes

Nicola Rosaia*

This version: July 8, 2021

Abstract

This paper studies estimation of undiscounted Markov decision processes (MDPs). Exploiting convex analytic methods, it argues that undiscounted MDPs can be treated as static discrete choice models over state-action frequencies, leveraging this idea to derive a conjugate duality framework for studying this type of models. It then exploits this framework to draw implications in several dimensions. First, it characterizes the empirical content of undiscounted MDPs, analyzing how exclusion or parametric restrictions can produce identification of agents' payoffs, and providing an axiomatic characterization of the undiscounted dynamic logit model; second, it proves convergence of simple inversion algorithms based on progressive Tâtonnements, and investigates novel estimation strategies based on these. Finally, it shows that the dual framework extends to models with persistent fixed effects and to models where certain actions or states are unobserved.

*Harvard University, Department of Economics, Littauer Center, Cambridge MA 02138, USA, nicolarosaia@g.harvard.edu
I am thankful to my advisors Myrto Kalouptsidi, Robin Lee, Ariel Pakes for constant advice and support, and to Alfred Galichon, Tomasz Strzalecki and Elie Tamer for helpful comments.

Introduction

Over the last few decades, Markov decision processes (MPDs) have established themselves as standard tools in economic analysis. The number of their economic applications has grown quickly, spanning from labor economics to industrial organization, public finance, health economics, and many others. A standard result is that, in these models, agents' discount factor is not identified from aggregate choice data.[1] Given this, the standard practice is to estimate MDPs by imposing a calibrated discount factor. In many situations, this is justified by the fact that the researcher has prior information regarding the discount factor. There are applications, however, where either this prior information is lacking, or there is indication that the discount factor is large. When this is the case, the standard approach is to impose a large but arbitrary discount factor, and to argue that a further increase should produce an insignificant improvement in the ability of the model to fit the data.[2] This paper takes an alternative route, by considering instead the problem of estimating undiscounted MDPs, that is, MDPs in which agents are assumed to not discount future payoffs.[3] In doing so, it shows how Convex Analytic methods can be fruitfully applied in studying this type of models, arguing that not only standard results on the estimation of discounted MDPs have simple analogues in the undiscounted case, but also that undiscounted models can sometimes be more convenient than their discounted counterparts. The reason being that, in undiscounted models, policy functions can be represented linearly in the payoff space by means of the induced state-action frequencies - the unconditional frequencies at which agents make different choices in the long-run. This implies that undiscounted MDPs can be treated as models of static discrete choice over state-action frequencies. The paper exploits this idea to derive a natural dual framework for studying this type of models, showing that: i) the gradient of the value of an undiscounted MDP provides a mapping from the state and choice-specific payoffs to the optimal state-action frequencies; and ii) the gradient of the convex conjugate of the value of an undiscounted MDP provides the inverse mapping, from state-action frequencies to the state and choice-specific payoffs inducing them as the outcome of agents' optimal behavior. In short, state-action

[1] See for instance Manski [1993], Rust [1994], Magnac and Thesmar [2002].

[2] For instance, Rust [1987] writes: "I was not able to precisely estimate the discount factor β... I did note a systematic tendency for the estimated value of β to be driven to 1... This suggests that if Harold Zurcher is actually minimizing long run average costs, an estimation algorithm based on discounted costs would use Abel's theorem and attempt to drive β to 1. This might be what's happening here".

[3] Undiscounted MDPs are widely popular tools in engineering and operations research. A textbook treatment can be found in Bertsekas [1995], while an extensive survey containing many references is given by Arapostathis et al. [1993]. To my knowledge, perhaps surprisingly, the problem of estimating these models has never been studied.

frequencies and agents' payoffs are related in the sense of conjugate duality.[4] This framework is then exploited to draw implications in several dimensions.

First, the empirical content of undiscounted MDPs is characterized. In particular, the paper shows that aggregate choice data identify the expected average payoffs associated with every stationary policy up to a constant. As a consequence, not only payoffs are not identified, but also their exact degree of under-identification can be determined. In order to identify the payoffs, additional restrictions must be imposed. However, identifying linear restrictions can be precisely characterized. In a nutshell, the paper shows that identification requires that two conditions are satisfied: i) the analyst must impose enough linear restrictions on the vector of payoffs that are linearly independent from the set of all possible state-action frequencies; and ii) the average payoff associated to some stationary policy must be normalized. In particular, these conditions relate the linear structure of exclusion or parametric restrictions to that of the state-transition probabilities of the MDP. While these identification results are comparable to those for discounted models, slightly stronger conditions are needed for identification in the undiscounted case.[5]

Second, duality implies that the problem of inverting an undiscounted MPD, namely that of finding a system of payoffs that would lead to the choice frequencies observed in the population, is a convex optimization. In this optimization, the objective function's gradient is given by the difference between the state-action frequencies observed in the population and those resulting from agents' optimizing behavior under a candidate system of payoffs. Hence, inversion can be performed by means of standard gradient-based methods, which correspond to simple Tâtonnement procedures in which the state and choice-specific payoffs are progressively increased proportionally to the difference between the observed state-action frequencies and those implied by the model. The paper shows that, under mild regularity conditions,

[4]This result is closely related to a well-know duality in static discrete choice models, where agents' payoffs and optimal choice probabilities are similarly related in terms of conjugate duality. This has been proven by Chiong et al. [2016], who use it to develop an estimator for discounted MDPs based on optimal transport methods, and recently exploited by Shi, Shum, and Song [2018] to derive identifying inequalities for static discrete choice models multinomial choice models with individual fixed effects. In dynamic models with discounting, this duality relates agents' optimal conditional choice probabilities with the optimal continuation values associated with different choices at every state, both of which are endogenous objects.

[5]Hence this paper relates to the literature on identification of discounted MDPs, such as Rust [1987], Magnac and Thesmar [2002], and the contributions that followed. In particular, the results on identification presented here have well-known analogues for discounted MDPs - one slight difference being that the latter usually deal with restrictions on the continuation values associated with different choices, while here I consider direct restrictions on the choice-specific payoffs. Most known results on identification of discounted MDPs are independent on the structure of state transitions - for instance, normalizing one payoff to zero at every state always yields identification. This is not the case for undiscounted models - for instance, normalizing one payoff to zero at every state might or might not yeld identification, depending on the state-transition probabilities.

these algorithms are guaranteed to converge monotonically to a solution. It then investigates novel estimation strategies which are based on these inversion procedures.

Following Rust [1987] and most of the subsequent literature, these results are presented for models satisfying the so-called Conditional Independence assumption, which rules out serially persistent forms of unobserved heterogeneity among agents. The dual framework can be exploited to characterize the restrictions imposed by this assumption on counterfactuals where alternatives are removed from agents' choice set at certain states. In particular, the paper defines a measure of statistical distance between state-action frequencies, and shows that Conditional Independence requires that the relative distance of any pair of state-action frequencies from the observed one is independent of the alternatives available to agents. This condition is a generalization of the well-known Independence of Irrelevant Alternatives [Luce, 1959], which characterizes the logit model in static discrete choice set-ups. Inter alia, this provides an axiomatic characterization of the undiscounted dynamic logit model.

The dual formulation of undiscounted MDPs makes it especially convenient to work with state-action frequencies as opposed to conditional choice probabilities. In particular, the paper concludes by showing that the main results extend directly to situations where the the analyst's observations are obtained by aggregating state-action frequencies in a linear fashion. It exemplifies this property by considering two instances of such linear aggregation. First, it considers mixed i.i.d. models, which arise when agents' payoffs are persistently heterogeneous, and the analyst observes a state-action frequency obtained by averaging state-action frequencies of different types of agents. Second, it considers the case in which the analyst cannot distinguish every action and state visited by agents, but instead observes the cumulative frequencies of a coarser partition of states and actions.

The paper is organized as follows: Section 1 introduces undiscounted MDPs and describes their optimal solutions; Section 2 introduces their conjugate duality framework; Section 3 describes their empirical content; Section 4 analyzes the identifying power of linear restrictions; Section 5 deals with inversion and estimation; Section 6 describes the main extensions. The proofs are presented in the Appendix.

1 Undiscounted Markov Decision Processes

1.1 Environment

This Section introduces the basic dynamic discrete choice setup. A state variable $x \in X$ can take a finite number of values. Agents choose actions $a \in A$ from a finite set. The per-period utility that an agent derives from choosing a is state x is

$$\boldsymbol{u}(a, x) + \epsilon(a)$$

where $\boldsymbol{u} \in \mathbb{R}^{A \times X}$ is a vector of structural choice-specific payoffs, and $\epsilon \in \mathbb{R}^{A}$ is a vector of utility shocks associated with each action, which is i.i.d. across agents and time periods. ϵ is distributed according to a distribution F which is assumed to be known to the analyst.

The pair (x, ϵ) evolves according to a first-order controlled Markov process. Conditional on the current state x and choice a, the future value of x, denoted by x', is drawn with probability $T(x'|a, x)$. The future value of ϵ, denoted by ϵ', is then drawn from F. In particular, following Rust [1987] and most of the subsequent literature, this rules out serially persistent heterogeneity across agents. That is, the so-called conditional independence assumption is imposed.

Assumption 1. *(Conditional Independence) Conditionally on the current state x, the distribution of ϵ is independent on the previous history of play:*

$$Pr(x', \epsilon'|x, \epsilon, a) = Pr(\epsilon'|x')Pr(x'|x, a)$$

Although the main results have an equivalent in the general case, for simplicity of exposition I assume that F is absolutely continuous and has full support. While this assumption covers all models commonly used in applications, it is also convenient since it allows to neglect tie-breaking in agents' optimization, and to obtain point identification results.

Assumption 2. *The distribution F is absolutely continuous and has full support*

In order to further reduce the expository noise, I confine the analysis to the case in which the state space is accessible, meaning every state can be reached in finite expected time from every other state by following an appropriate sequence of choices. This assumption can be stated formally as follows.

Assumption 3. *(Accessibility) For every set of states $Y \subsetneq X$, there exist $y \in Y$, $x \in X \setminus Y$ and $a \in A$ such that $T(x|a, y) > 0$.*

In other words, say that y is accessible from x, written $x \to y$, if there exist sequences of states $x_1, ..., x_{n+1}$ and actions $a_1, ..., a_n$ such that $x_1 = x$, $x_{n+1} = y$ and $T(x_{k+1}|a_k, x_k) > 0$ for every $k = 1, .., n$. Accessibility requires that the whole state space is irreducible according to the relation $\to$. If this was not the case, the results of the paper would still apply separately to every irreducible class of states. On the one hand, states not belonging to any irreducible class are irrelevant for the analysis since they are never visited in the long run, no matter which choices the agents make. On the other hand, Accessibility does impose that there are no terminal (i.e. absorbing) states. Hence, in a sense, it limits the scope of the analysis to truly infinite-horizon models. This limitation is natural, however, since undiscounted models with terminal states are formally closer to models with discounting. Indeed, assumptions of this kind are commonly imposed in general treatments of undiscounted MDPs.[6]

1.1.1 Example: Rust's engine replacement model

This Section introduces a classic examples of such an environment, which I will later use to illustrate the main results. Rust [1987] studied a model in which Harold Zurcher, the owner of a fleet of buses, makes optimal dynamic engine replacement decisions. The state of a bus is $x \in \{x_1, ..., x_N\}$, representing the mileage of its engine, with $x_1 < ... < x_N$. It is assumed that each bus incurs an operating cost $c(x)$ which depends on the current mileage. The engine of a bus can be replaced at a cost RC which is mileage-independent, in which case the mileage is reset to its lowest possible value x_1 with certainty. If the engine is not replaced, the mileage does not decrease, and increases to with some probability. This model is decribed by the following primitives: $A = \{r, nr\}$, $T(x_1|r, x_n) = 1$ for all n, $T(x_{n'}|nr, x_n) = 0$ for all $n' < n$, $T(x_{n+1}|nr, x_n) > 0$ for all $n < N$, $u(nr, x) = -c(x)$ and $u(r, x) = -c(x) - RC$ for all x. Notice that Assumption 3 holds since x_1 is reached with probability one from every state if the engine is replaced, and each state can be reached with positive probability from x_1 by never replacing the engine.

[6] See for instance Bertsekas [1995, Chapter 4]

1.1.2 Notation and terminology

At this point it is useful to introduce some notation and terminology that will be used throughout the paper.

Linear algebra

Vectors and matrices For a generic set S, I will often treat elements of $\mathbb{R}^S$ as vectors in $\mathbb{R}^{|S|}$, where $|S|$ denotes the size of S. If $N > 0$ and $C \subseteq \mathbb{R}^N$ is a set of vectors in $\mathbb{R}^N$, abusing notation, C will also denote the matrix whose columns are the elements of C, and $|C|$ will denote the number of its elements (or columns).

Linear operations The usual linear algebra notation will apply. So, for instance, if b and c are conformable vectors, $b \cdot c$ will denote their inner product, and $b = c$ (resp. $b \geq c$) will mean that every coordinate of b is equal to (resp. greater than) the respective coordinate of c. C' will denote the transpose of matrix C, and for any conformable matrix B, CB will denote their matrix product.

Linear combinations Standard notation and terminology will apply. For instance, I will say that a set of vectors (or matrix) C is linearly independent if, for every conformable vector α, $C\alpha = 0$ if and only if $\alpha = 0$. $\dim C$ will denote the dimension of C, namely the size of a maximal linearly independent subset (a linear basis) of C. When C is finite, $\dim C$ equals the rank of matrix C, denoted by $\text{rank}C$. Finally, $\text{Span}C$ will denote the vector space spanned by the columns of C.

Notable vectors and matrices When necessary to avoid confusion, subscripts will denote the size of vectors and the number of rows and columns of matrices. So, in particular, 0_N and 1_N will denote vectors with N coordinates equal to zero and N coordinates equal to one, respectively, while $0_{N\times M}$ and $1_{N\times M}$ will denote matrices in $\mathbb{R}^{N\times M}$ with entries all equal to zero and entries all equal to one, respectively. The notation $1\{\cdot\}$ will be used for indicators. So, for instance, if $1\{a, x\}$ will denote the vector in $\mathbb{R}^{|A||X|}$ whose coordinates are all null except for a unit entry at the a, x-th coordinate.

Other notation

Probabilities I will use the symbol Δ to denote probability simplices. So, in particular, $\Delta A \subseteq \mathbb{R}^{|A|}$ and $\Delta(A \times X) \subseteq \mathbb{R}^{|A||X|}$ will denote the sets of probability measures over A and $A \times X$, respectively. If $\boldsymbol{\mu} \in \Delta(A \times X)$, I will denote by $\boldsymbol{\mu}_X \in \Delta X$ its marginal over states, defined by $\boldsymbol{\mu}_X(x) = \sum_a \boldsymbol{\mu}(a, x)$ for all $x \in X$.

Gradients I will use the symbol ∇ to denote gradients and sub-gradients of convex functions.[7] Formally, for an arbitrary convex function $g : \mathbb{R}^N \to \mathbb{R}$, $\nabla g(y) \subseteq \mathbb{R}^N$ will denote its sub-gradient evaluated at $y \in \mathbb{R}^N$, and if g is differentiable at y then, abusing notation, $\nabla g(y) \in \mathbb{R}^N$ will denote its gradient at y.

Bold symbols The paper will sometimes draw comparisons between undiscounted MDPs and static discrete choice models. To avoid confusion, when dynamic objects have a static analogue, the same symbol will be used for both objects, but dynamic objects will be denoted in bold. So, for instance, $u \in \mathbb{R}^{|A|}$ will denote a vector of payoffs in a static choice, while $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ will denote a vector of state and choice-specific payoffs. Similarly, $\sigma \in \Delta A$ will denote a vector of static discrete choice probabilities, while $\boldsymbol{\sigma} \in (\Delta A)^X$ will denote a dynamic system of conditional choice probabilities. Other symbols will be introduced later in the text.

1.2 Optimality

Given these preliminaries, this Section describes the solution of the undiscounted MDP, meaning the behavior of forward-looking agents who maximize the expected long-run average of their future rewards. Formally, agents choose a *stationary policy* $\boldsymbol{\pi}$, which associates an action a to each state x and realization of the i.i.d. shock ϵ:

$$\boldsymbol{\pi} : X \times \mathbb{R}^A \to A$$

Their goal is to maximize the *expected average payoff* from every initial state x_0, which is defined by

$$\boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi}, x_0) \equiv \lim_{T \to \infty} \frac{1}{T+1} \mathrm{E}_{\boldsymbol{\pi},F}[\sum_{t=0}^{T} (\boldsymbol{u}(a_t, x_t) + \epsilon_t(a_t))|x_0]$$

[7] See Section A for a definition of sub-gradients.

The expectation above is taken with respect to the distribution of future states, actions and shocks induced by $\boldsymbol{\pi}$, under the distribution F of the i.i.d. shocks and the transition kernel T, starting from state x_0. That is, conditional on x_t, ϵ_t is drawn according to F, and $a_t = \boldsymbol{\pi}(x_t, \epsilon_t)$. Then, conditional on x_t and a_t, x_{t+1} is drawn according to $T(a_t, x_t)$.

Definition 1. A stationary policy $\boldsymbol{\pi}$ is *optimal* under $\boldsymbol{u}$ if it solves

$$\max_{\boldsymbol{\pi}'} \boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi}', x) \tag{1}$$

for every initial state x.

For any policy $\boldsymbol{\pi}$ and initial state x_0, define the long-run *state-action frequency* $\boldsymbol{\mu}(\boldsymbol{\pi}, x_0) \in \Delta(A \times X)$ under the stochastic process generated by $\boldsymbol{\pi}$ from x_0:

$$\boldsymbol{\mu}(a, x|\boldsymbol{\pi}, x_0) = \lim_{T\to\infty} \frac{1}{T+1} \mathrm{E}_{\boldsymbol{\pi},F}[\sum_{t=0}^{T} 1\{a_t = a, x_t = x\}|x_0] \text{ for every } a, x$$

Equivalently, $\boldsymbol{\mu}(\boldsymbol{\pi}, x_0)$ can be obtained from the system of *conditional choice probabilities* generated by $\boldsymbol{\pi}$. To see this, for every a, x, let

$$\boldsymbol{\sigma}(a, x|\boldsymbol{\pi}) \equiv \Pr_F[\boldsymbol{\pi}(x, \epsilon) = a] \equiv \int 1\{\boldsymbol{\pi}(x, \epsilon) = a\} \mathrm{d}F$$

be the probability according to which $\boldsymbol{\pi}$ selects action a in state x. $\boldsymbol{\sigma}(\boldsymbol{\pi})$ generates a system of Markov state-transitions, where the probability of transitioning from state x to state x' is given by $\sum_a T(x'|a, x)\boldsymbol{\sigma}(a, x|\boldsymbol{\pi})$. Then $\boldsymbol{\mu}_X(\boldsymbol{\pi}, x_0)$ - the marginal of $\boldsymbol{\mu}(\boldsymbol{\pi}, x_0)$ over states - is the stationary distribution over states associated with this Markov chain with initial state x_0, and $\boldsymbol{\mu}(a, x|\boldsymbol{\pi}, x_0) = \boldsymbol{\mu}_X(x|\boldsymbol{\pi}, x_0)\boldsymbol{\sigma}(a, x|\boldsymbol{\pi})$ for every a, x. Notice that we can write

$$\boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi}, x_0) = \boldsymbol{\mu}(\boldsymbol{\pi}, x_0) \cdot \boldsymbol{u} + \sum_x \boldsymbol{\mu}_X(x|\boldsymbol{\pi}, x_0)\mathrm{E}_F[\epsilon(\boldsymbol{\pi}(x, \epsilon))]$$

In words, the expected average payoff of $\boldsymbol{\pi}$ can be broken down in two separate components: the expected average structural payoff, and the expected average utility coming from the idiosyncratic shocks. The structural component is represented linearly by the state-action frequency $\boldsymbol{\mu}(\boldsymbol{\pi}, x_0)$.

It is important to notice that, under Assumption 2, an optimal policy $\boldsymbol{\pi}$ must induce full-support choice probabilities. That is, we must have

$$\boldsymbol{\sigma}(a, x|\boldsymbol{\pi}) > 0 \text{ for every } a, x \tag{2}$$

Hence Accessibility implies that every state is recurrent under the Markov state-transitions induced by $\boldsymbol{\sigma}(\boldsymbol{\pi})$. This implies that the long-run state-action frequency associated with $\boldsymbol{\pi}$ does not depend on the initial state x_0, so that we can define $\boldsymbol{\mu}(\boldsymbol{\pi}) \equiv \boldsymbol{\mu}(\boldsymbol{\pi}, x_0)$ and $\boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi}) \equiv \boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi}, x_0)$ for an arbitrary state x_0. This allows to define the optimal state-action frequency and the value of the undiscounted MDP as follows.

Definition 2. $\boldsymbol{\mu} \in \Delta(A \times X)$ is *optimal under* $\boldsymbol{u}$, written $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{u})$, if $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\pi})$ for some optimal policy $\boldsymbol{\pi}$. The *value* $\boldsymbol{w}(\boldsymbol{u})$ of the undiscounted MDP under $\boldsymbol{u}$ is the expected average payoff $\boldsymbol{w}(\boldsymbol{u}|\boldsymbol{\pi})$ achieved by some optimal policy $\boldsymbol{\pi}$.

If $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{u})$ I will also say that $\boldsymbol{\mu}$ is *rationalized* by $\boldsymbol{u}$, to emphasize the fact $\boldsymbol{\mu}$ is consistent with the behavior of optimizing agents under some vector of choice-specific payoffs. Similarly, I will often say that $\boldsymbol{\sigma} \in (\Delta A)^X$ is optimal under (or rationalized by) $\boldsymbol{u}$, if $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\pi})$ for some optimal policy $\boldsymbol{\pi}$. In what follows, I will focus on optimal state-action frequencies and conditional choice probabilities, leaving policies in the background. In doing so, I take the perspective of an analyst who does not observe agents' i.i.d. shocks, but only their actions and states. For this reason, and since the analyst's observations can be equivalently represented by means of either state-action frequencies or conditional choice probabilities, I will often generically refer to both as *choice outcomes*.

1.3 Recursive formulation and relative value iteration

This Section shows that optimal choice outcomes are characterized by a set of recursive equations analogous to Bellman equations in discounted models, which can be computed by means of backward-induction algorithms. Most results of this Section are well known, and the goal is to provide a self-contained introduction to undiscounted MDPs. However, proofs are provided in the Appendix for completeness.

1.3.1 Recursive equations

First, as already noted in previous Section, Accessibility implies that the value of the MDP is independent of the initial state. It also implies that its solution is characterized by the following set of recursive equations.

Proposition 1. *$\boldsymbol{\sigma}$ is optimal under $\boldsymbol{u}$ if and only if there exist $V \in \mathbb{R}^X$ and $\boldsymbol{w} \in \mathbb{R}$ such that*

$$V(x) + \boldsymbol{w} = E_F \max_a [\boldsymbol{u}(a,x) + T(a,x) \cdot V + \epsilon(a)] \text{ for every } a, x \tag{3}$$

and

$$\boldsymbol{\sigma}(a,x) = Pr_F[a \in \arg\max_{a' \in A} [\boldsymbol{u}(a',x) + T(a',x) \cdot V + \epsilon(a')]] \text{ for every } a, x \tag{4}$$

Moreover, $\boldsymbol{w} = \boldsymbol{w}(\boldsymbol{u})$ if and only if there exists V such that $\boldsymbol{w}$ and V satisfy 3.

The Equations in 3 can be thought as the undiscounted analogues of Bellman equations. Intuitively, V can be thought as a vector of relative valuations attached to different states. Indeed, notice that only the differences between different coordinates of V are pinned down by Condition 3 - that is, adding or subtracting a constant from all the coordinates of V would leave the condition satisfied. Hence, in particular, one coordinate of V can always be normalized to zero. The quantities $\boldsymbol{u}(a,x)+T(a,x)\cdot V+\epsilon(a)$ can be thought as the relative continuation values associated with different actions in different states, under a given realization ϵ of the utility shock. The Equations in 4 relate these relative valuations to the optimal conditional choice probabilities.

1.3.2 Undiscounted limit

To gain further intuition, It can be shown that the Equations in 3 can be obtained as the limits of Bellman equations as the discount factor goes to one. When agents discount future payoffs according to a discount factor β, the optimal system $\boldsymbol{\sigma}^\beta$ of choice probabilities is defined by

$$\boldsymbol{\sigma}^\beta(a,x) = \Pr_F[a \in \arg\max_{a' \in A} [\boldsymbol{u}(a',x) + \beta T(a',x) \cdot V^\beta + \epsilon(a')]] \text{ for every } a, x \tag{5}$$

where $V^\beta \in \mathbb{R}^X$ is the value function defined recursively by the Bellman equations

$$V^\beta(x) = \mathrm{E}_F[\max_{a\in A}[\boldsymbol{u}(a,x) + \beta T(a,x)\cdot V^\beta + \epsilon(a)]] \text{ for every } x \ .$$

Proposition 2. *Let $\boldsymbol{\sigma}, V, \boldsymbol{w}$ be as in Proposition 1. Then, as $\beta \to 1$, $\boldsymbol{\sigma}^\beta \to \boldsymbol{\sigma}$ and*

$$(1-\beta)V^\beta(x) \to \boldsymbol{w} \text{ and } V^\beta(x) - V^\beta(x') \to V(x) - V(x') \text{ for every } x, x' \ .$$

For the case without shocks, this result was first proven by Blackwell [1962], who also showed that there exist policies which are optimal for every β above a certain threshold. The existence of such policies then implies, in the case without shocks, that every limit of discounted solutions is also an undiscounted solution. On the other hand, the converse is true only under full support shocks, since in this case the solutions of both the discounted and undiscounted versions of the problem can be shown to be unique.

Intuitively, the differences between the valuations attached to different states grow at similar rates as the discount factor increases, due to the fact that every state can be reached in finite expected time from any other state (by Accessibility). This implies that the limit of the average expected payoff $(1-\beta)V^\beta(x)$ is independent of the state x. This limit is the optimal expected average per-period payoff associated with the undiscounted problem. Hence one can take the no-discounting limit of the Bellman equations by re-writing them in relative terms:

$$\underbrace{V^\beta(x) - V^\beta(x')}_{\to V(x)-V(x')} + \underbrace{(1-\beta)V^\beta(x')}_{\to \boldsymbol{w}} = \mathrm{E}_F[\max_{a\in A} \boldsymbol{u}(a,x) + \beta T(a,x)\cdot[\underbrace{V^\beta - V^\beta(x')}_{\to V - V(x')}] + \epsilon(a)] \tag{6}$$

which yields the recursive equations in 3. Similarly, taking the limit of the optimal choice probabilities yields the undiscounted ones.

1.3.3 Computation

Given this recursive formulation, a natural candidate algorithm for solving the undiscounted MDP is to generate successively the finite-horizon continuation values V^k, $k = 0, 1, 2, ...$ starting with some initial

vector V^0, where

$$V^{k+1}(x) = \mathrm{E}_F[\max_{a \in A} \boldsymbol{u}(a,x) + T(a,x)V^k + \epsilon(a)]$$

for every $k \geq 1$ and state x. It is then natural to speculate that, for every state x, the quantity $(1/k)V^k(x)$ converges to the optimal value $\boldsymbol{w}$ as $k \to \infty$. This procedure has two drawbacks. First, the calculation is impractical since the components of V^k typically diverge. Second, a corresponding vector of relative valuations V is not obtained. One may attempt to address these issues by subtracting some same scalar w^k from all components of V^k so that the differences $V^k(x) - w^k$ remain bounded. This suggests an algorithm as follows. Fix an arbitrary state $x_0 \in X$, and define the relative value operator $\Gamma : \mathbb{R}^X \to \mathbb{R}^X$ by

$$\Gamma V(x) = \mathrm{E}_F[\max_{a \in A} \boldsymbol{u}(a,x) + T(a,x) \cdot V + \epsilon(a)] - \mathrm{E}_F[\max_{a \in A} \boldsymbol{u}(a,x_0) + T(a,x_0) \cdot V + \epsilon(a)] \text{ for every } x.$$

For any arbitrary initial condition V^0, the sequence of iterations $V^{k+1} = \Gamma V^k$ defines the well know relative value iteration algorithm for solving undiscounted problems. Notice that, if the sequence $(V^k)_{k \geq 0}$ converges to some fixed point V^* of Γ then, letting $w^* = \mathrm{E}_F[\max_{a \in A} \boldsymbol{u}(a,x_0) + T(a,x_0) \cdot V^* + \epsilon(a)]$, the pair V^*, w^* satisfies the recursive Equations in 3. This algorithm was first introduced by White [1963], who also showed convergence under restrictive conditions. Convergence under slightly weaker conditions was later shown by Platzman [1977], while Federgruen et al. [1978] established necessary and sufficient conditions for the operator Γ to reduce to a contraction mapping, in which case relative value-iteration method exhibits a uniform geometric convergence rate. Convergence of this algorithm, however, is not guaranteed under the general conditions of this paper. I conclude this Section by showing that, on the other hand, adding a dampening step to this procedure always guarantees convergence. Formally, fix for some arbitrary $\alpha \in (0,1)$, and define the operator $\Gamma^\alpha : \mathbb{R}^X \to \mathbb{R}^X$ by

$$\Gamma^\alpha V(x) = \alpha V(x) + (1-\alpha)\Gamma V(x) \text{ for every } x \tag{7}$$

Notice that V^* is a fixed point of Γ^α if and only if it is a fixed point of Γ. The next Proposition shows that the iterates generated by Γ^α always converge to a fixed point.

Proposition 3. *Fix an arbitrary state $x_0 \in X$ and scalar $\alpha \in (0,1)$, and define the operator Γ^α as above. Γ^α has a unique fixed point V^*. Moreover, for any initial condition $V^0 \in \mathbb{R}^X$, the sequence $(V^k)_{k=0}^\infty$ defined by $V^{k+1} = \Gamma^\alpha V^k$ for all $k \geq 0$ converges to V^*.*

In words, adding a dampening step to relative value iteration provides an algorithm that always converges to a solution of the undiscounted MDP. To my knowledge, this has never been noted before. The proof relies on the well-known fact that the relative value operator Γ is non-expansive with respect to the so-called span semi-norm - see Bertsekas [1995, Chapter 4], and Appendix B.4 - and exploits a result by Ishikawa [1976] establishing convergence of the iterates generated as in 7 for non-expansive mappings.[8]

2 Duality

This Section presents a result that is central to the paper, showing that optimal state-action frequencies and choice-specific payoffs are related in the sense of conjugate duality. In order to provide intuition and background, I start by introducing the analogue of this result for static discrete choice models. I then move to the dynamic case, highlighting similarities and differences with the static set-up. Recall that, in order to distinguish static and dynamic objects, bold symbols are used for the latter. So, in particular, $\boldsymbol{\sigma} \in (\Delta A)^X$ will denote a system of conditional choice probabilities, while $\sigma \in \Delta A$ will denote a vector of choice probabilities; $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ will denote a vector of state and choice-specific payoffs, while $u \in \mathbb{R}^A$ will denote a vector of static payoffs associated to different actions; $\boldsymbol{w}$ will denote the value of an undiscounted MDP, while w will denote the value of a static discrete choice problem.

2.1 Duality in static discrete choice

Abstracting for a moment from dynamic considerations, consider a decision maker facing a static discrete choice among the alternatives in A, and suppose that his payoff for choosing action a is random and given by $u(a) + \epsilon(a)$, where $u \in \mathbb{R}^A$ and ϵ is distributed according to F. If $\sigma \in \Delta A$ is a vector of choice

[8] Actually it can be shown that there exists an integer k such that the k-th iterate of Γ always contracts, but its Lipschitz constant cannot be bounded away from 1, hence the need of exploiting results on non-expansive mappings.

probabilities, say that u *rationalizes* σ if

$$\sigma(a) = \Pr_F[u(a) + \epsilon(a) = \max_{a' \in A}[u(a') + \epsilon(a')]] \text{ for every } a \in A.$$

Define the ex-ante inclusive value of the decision problem as a function of u by

$$w(u) = \mathrm{E}_F \max_{a \in A}[u(a) + \epsilon(a)] \text{ for every } u \in \mathbb{R}^A$$

Define also its *convex conjugate* by

$$w^*(\sigma) = \max_{u \in \mathbb{R}^A}[\sigma \cdot u - w(u)] \text{ for every } \sigma \in \Delta A \ . \tag{8}$$

Example. When F belongs to the family of Logit distributions, w and w^* can be obtained in closed form. For instance, if F is the standard logit distribution, we have

$$w(u) = \log \sum_{a \in A} \exp u(a) + \gamma \text{ for every } u \in \mathbb{R}^A$$

and

$$w^*(\sigma) = \begin{cases} \sum_a \sigma(a) \log \sigma(a) - \gamma & \text{if } \sigma \in \Delta A \\ +\infty & \text{otherwise} \end{cases}$$

where γ is the Euler's constant. In words, $-w^*(\sigma)$ equals the entropy of σ up to a constant. For this reason, in what follows I will refer to the convex conjugate w^* under a generic distribution F as its *generalized entropy.*

The following result, first proven by Chiong et al. [2016], is a consequence of Fenchel's duality Theorem for static discrete choice models.[9]

Theorem 1. *The following are equivalent:*

i) u rationalizes σ

[9] See Rockafellar [1970, Section 31]

ii) σ solves

$$w(u) = \max_{\sigma \in \mathbb{R}^A} [\sigma \cdot u - w^*(\sigma)] \tag{9}$$

iii) u solves Problem 8

The idea behind this result is very simple. Intuitively, the quantity $-w^*(\sigma)$ can be thought as the maximum utility from shocks that is achievable by matching shocks to actions in a way that is consistent with σ. That is, the following result holds, which was proven by Galichon and Salanié [2020].

Proposition 4. *For $\sigma \in \Delta A$ we have*

$$-w^*(\sigma) = \max_{\pi \colon \mathbb{R}^A \to A} E_F[\epsilon(\pi(\epsilon))] \text{ s.t. } Pr_F[\pi(\epsilon) = a] = \sigma(a) \text{ for all } a \in A$$

where $E_F[\epsilon(\pi(\epsilon))]$ the expectation with respect to the realization of ϵ taken according to F.

Hence, a static discrete choice problem can be split in two nested problems: an outer problem in which an optimal vector of choice probabilities is chosen, and an inner problem - with value $-w^*(\sigma)$ - in which an optimal matching between shocks and actions is chosen in order to maximize the expected utility arising from shocks, subject to it being consistent with σ.

w and w^* are both convex functions, ans the solutions of Problems 8 and 9 are characterized by their first order conditions.[10] It follows that u rationalizes σ if and only if $\sigma \in \nabla w(u)$. This result, known as the Williams–Daly–Zachary Theorem, characterizes the optimal choice probabilities in a discrete-choice model. It has been expounded in McFadden [1978] and Rust [1994], among others. On the other hand, it also holds that u rationalizes σ if and only if $u \in \nabla w^*(\sigma)$. This solves the identification problem, namely to determine the set of u that would lead to a given vector of choice probabilities σ. In the dynamic case, this provides a mapping from optimal choice outcomes to the optimal continuation values - in undiscounted models, to the optimal relative continuation values. That is, if V and $\boldsymbol{w}$ are as in Proposition 1, it follows that a vector $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ of choice-specific payoffs rationalizes a system $\boldsymbol{\sigma}$ of conditional choice probabilities if and only if

$$(\boldsymbol{u}(a,x) + T(a,x) \cdot V)_{a \in A} \in \nabla w^*(\boldsymbol{\sigma}(x)) \text{ for every } x.$$

[10] See Appendix A

Chiong et al. [2016] exploit this idea - and Proposition 4 - in order to develop an estimator for discounted MDPs based on optimal transport methods. In contrast, the next Section shows that, in undiscounted models, duality directly relates the state and choice-specific payoffs to the optimal state-action frequencies.

2.2 Duality in undiscounted MDPs

The first step is to define the generalized entropy of state-action frequencies. To do so, for every measure $\boldsymbol{\mu} \in \mathbb{R}^{A \times X}$ such that $\boldsymbol{\mu} \geq 0$, not necessarily in $\Delta(A \times X)$, let $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$ be an arbitrary system of conditional choice probabilities satisfying

$$\boldsymbol{\mu}(a,x) = \boldsymbol{\mu}_X(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) \text{ for every } a, x \ .$$

That is, at any state x, if $\boldsymbol{\mu}_X(x) > 0$ then $\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) \equiv (\boldsymbol{\mu}(a,x)/\boldsymbol{\mu}_X(x))_{a \in A}$ is the vector of choice probabilities induced by $\boldsymbol{\mu}$ at x, while if $\boldsymbol{\mu}_X(x) = 0$ I let $\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)$ be an arbitrary vector of choice probabilities. Given this, the *generalized entropy* $\boldsymbol{w}^*(\boldsymbol{\mu})$ *of* $\boldsymbol{\mu}$ is defined as the $\boldsymbol{\mu}_X$-weighted sum of the generalized entropies of $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$ at each state:

$$\boldsymbol{w}^*(\boldsymbol{\mu}) = \sum_x \boldsymbol{\mu}_X(x) w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) \text{ for every } x.$$

Hence $\boldsymbol{w}^*$ is well-defined for every $\boldsymbol{\mu} \geq 0$, and so in particular for every $\boldsymbol{\mu} \in \Delta(A \times X)$.

Example. When F belongs to the family of Logit distributions, $\boldsymbol{w}^*$ can be obtained in closed form. For instance, if F is the standard logit distribution, we have

$$\boldsymbol{w}^*(\boldsymbol{\mu}) = \sum_{a,x} \boldsymbol{\mu}(a,x) \log \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) - \gamma$$

The second step is to define the set of all state-action frequencies $\boldsymbol{\mu} \in \Delta(A \times X)$ that can arise under some stationary policy from some initial state - the set of frequencies that can in principle be observed by the analyst. For this to be the case, since any system of conditional choice probabilities can be generated by some stationary policy, the only restriction is that $\boldsymbol{\mu}_X$ be a stationary distribution over

states associated with the Markov state-transitions induced by $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$. That is, $\boldsymbol{\mu}$ must satisfy

$$\boldsymbol{\mu}_X(x) = \sum_{x'} \boldsymbol{\mu}_X(x') \sum_{a'} \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a', x') T(x|a', x') \text{ for all } a, x \ .$$

This, together with the definition of $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$, yields the following.

Definition 3. A state-action frequency $\boldsymbol{\mu} \in \Delta(A \times X)$ is *stationary* if it satisfies

$$\sum_{a} \boldsymbol{\mu}(a, x) = \sum_{a', x'} \boldsymbol{\mu}(a', x') T(x|a', x') \text{ for all } x.$$

The set of all stationary state-action frequencies is denoted by M.

Given these preliminaries, duality extends to the dynamic case as follows.

Theorem 2. *$\boldsymbol{w}$ and $\boldsymbol{w}^*$ are both convex and continuously differentiable, and $\boldsymbol{w}^*$ is strictly convex. Moreover, the following are equivalent:*

i) $\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$

ii) $\boldsymbol{\mu}$ solves

$$\boldsymbol{w}(\boldsymbol{u}) = \max_{\boldsymbol{\mu} \in M} [\boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu})] \tag{10}$$

iii) $\boldsymbol{u}$ solves

$$\boldsymbol{w}^*(\boldsymbol{\mu}) = \max_{\boldsymbol{u} \in \mathbb{R}^{|A||X|}} [\boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u})] \tag{11}$$

In analogy with the static case, the quantity $-\boldsymbol{w}^*(\boldsymbol{\mu})$ can be thought as the maximum expected average utility from shocks that is achievable by matching shocks to actions in each state in a way that is consistent with $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$. Hence, the problem of choosing an optimal stationary policy can be split into two nested problems: an outer problem in which an optimal state-action frequency is chosen, and an inner problem - with value $-\boldsymbol{w}^*(\boldsymbol{\mu})$ - in which an optimal stationary policy is chosen in order to maximize the expected utility arising from shocks, subject to consistency with $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$.

In the case without shocks, it is well known that undiscounted MDPs admit a linear programming representation.[11] ii) extends this result to models with idiosyncratic shocks. It shows that finding an

[11] See Kallenberg [1994b] and Kallenberg [1994a] for a survey of linear programming methods for solving undiscounted MDPs.

optimal state-action frequency is equivalent to solving a particular type of convex optimization problem: a regularized linear optimization, with regularization given by $\boldsymbol{w}^*$. Hence, in alternative to the methods discussed in Section 1.3, convex programming algorithms can be used for computation. Moreover, in analogy with the static case, this implies that $\boldsymbol{\mu}(\boldsymbol{u}) = \nabla \boldsymbol{w}(\boldsymbol{u})$, which can be thought as a version of Roy's identity for undiscounted MDPs. Conversely, iii) implies that finding a vector of state and choice-specific payoffs rationalizing a given state-action frequency is formally equivalent to solving a convex optimization problem. However, this result does not yet determine the set of $\boldsymbol{u}$ that would lead to a given stationary measure $\boldsymbol{\mu}$. That is, it does not solve the identification problem, which will be the focus Section 4.

Before concluding this Section it is worth making two additional remarks. First, in order to be rationalizable by some vector of choice-specific payoffs, a state-action frequency $\boldsymbol{\mu}$ must have *full support*. That is, it must be such that $\boldsymbol{\mu}(a, x) > 0$ for every a, x, since any optimal stationary policy induces full-support choice probabilities. This implies that the constraint $\boldsymbol{\mu} \geq 0$ in Problem 10 never binds. In what follows, I will denote by M_+ the set of stationary state-action frequencies of full support. Second, in analogy with the static case, convex conjugacy between $\boldsymbol{w}$ and $\boldsymbol{w}^*$ still holds, provided that the domain of $\boldsymbol{w}^*$ is restricted to M.[12]

3 Empirical content of undiscounted MDPs

Notice that Problem 10 depends on $\boldsymbol{u}$ only through the set of average payoffs

$$\{\boldsymbol{\nu} \cdot \boldsymbol{u} : \quad \boldsymbol{\nu} \in M\} \ . \tag{12}$$

Therefore, different vectors $\boldsymbol{u}$ yielding the same set of average payoffs cannot be told apart from the observation of $\boldsymbol{\mu}$. Moreover, adding or subtracting a constant from all these averages does not affect the problem's solution, hence we can at most hope to identify the set 12 up to a constant. The next proposition shows that this is indeed the empirical content of undiscounted MDPs.

[12] Formally, defining

$$\tilde{\boldsymbol{w}}^*(\boldsymbol{\mu}) = \begin{cases} \boldsymbol{w}^*(\boldsymbol{\mu}) & \text{if } \boldsymbol{\mu} \in M \\ +\infty & \text{otherwise} \end{cases}$$

for every $\boldsymbol{\mu} \in \mathbb{R}^{A \times X}$, $\tilde{\boldsymbol{w}}^*$ and $\boldsymbol{w}$ are convex conjugates. Hence, similarly to the static case, $\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$ if and only if $\boldsymbol{u} \in \nabla \tilde{\boldsymbol{w}}^*(\boldsymbol{\mu})$.

Proposition 5. *$\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$ if and only if there exists a scalar $k \in \mathbb{R}$ such that*

$$\boldsymbol{\nu} \cdot \boldsymbol{u} = \boldsymbol{\nu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}) + k \text{ for every } \boldsymbol{\nu} \in M \ . \tag{13}$$

Moreover, if $\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$, then k satisfies Condition 13 if and only if $k = \boldsymbol{w}(\boldsymbol{u})$.

In words, $\boldsymbol{\mu}$ identifies the expected average payoff of each policy, up to a constant. The intuition behind this result is that undiscounted MDPs can be seen as models of static choice among stationary policies. In particular, they can be seen as models of static discrete choice among a finite subset of policies spanning the whole payoff space. To see this, let B be an arbitrary linear basis of M. That is, B is such that, for every $\boldsymbol{\mu} \in M$, there is a unique $\alpha \in \mathbb{R}^B$ such that $\sum_{\boldsymbol{\nu} \in B} \alpha(\boldsymbol{\nu}) = 1$ and $\boldsymbol{\mu} = B\alpha$. It is easy to see that the undiscounted MDP can be seen as a static discrete choice over B. Indeed, letting $\boldsymbol{\upsilon} \equiv (\boldsymbol{\nu} \cdot \boldsymbol{u})_{\boldsymbol{\nu} \in B} \in \mathbb{R}^B$ be the vector whose coordinates are the expected average payoffs associated with the elements of B, Problem 10 is equivalent to

$$\max_{\alpha \in \mathbb{R}^B} [\alpha \cdot \boldsymbol{\upsilon} - \boldsymbol{w}^*(B\alpha)] \text{ s.t. } \sum_{\boldsymbol{\nu} \in B} \alpha(\boldsymbol{\nu}) = 1 \text{ and } B\alpha \geq 0 \tag{14}$$

In particular, since only state-action frequencies in M_+ can be rationalized, the non-negativity constraint never binds. So Problems 9 and 14 have the same constraints, and the only difference between them is the structure of the generalized entropies.[13] This clarifies the sense in which undiscounted MDPs are essentially models of static discrete choice over B. Notice that the optimality conditions of Problem 14 can be written as

$$\exists\, k \in \mathbb{R} : \ \boldsymbol{\nu} \cdot \boldsymbol{u} = \boldsymbol{\nu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}) + k \text{ for every } \boldsymbol{\nu} \in B \ . \tag{15}$$

Since B spans M, this is equivalent to Condition 13.

Example. When F belongs to the family of Logit distributions, $\nabla \boldsymbol{w}^*$ can be obtained in closed form.

[13] One difference that should be noted in particular is that, due to the structure of generalized entropies in models with idiosyncratic shocks, these models do not satisfy gross substitutes. The latter is a key property of static discrete choice models, yielding invertibility of a large class of demand models (see Berry et al. [1995] and Berry et al. [2013]). In contrast, idiosyncratic shocks typically induce complementarities between state-action frequencies. However, undiscounted MDPs preserve invertibility, and Section 5.1 shows that this be performed by means of simple Tâtonnement algorithms that exploit duality instead of gross substitutes.

For instance, if F is the standard logit distribution, we have

$$\frac{\mathrm{d}\boldsymbol{w}^*(\boldsymbol{\mu})}{\mathrm{d}\boldsymbol{\mu}(a,x)} = \log \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x).$$

Hence $\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$ if and only if there exists a scalar $k \in \mathbb{R}$ such that

$$\sum_{a,x} \boldsymbol{\nu}(a,x)\boldsymbol{u}(a,x) = \sum_{a,x} \boldsymbol{\nu}(a,x) \log \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) + k \text{ for every } \boldsymbol{\nu} \in M \ .$$

In general, the gradient $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ of the generalized entropy cannot be obtained in closed form. However Section 5.1 shows that, if F satisfies a mild regularity condition, $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ can be computed through simple algorithms.

Before concluding this Section, it is useful to note that that there is a simple way to construct such a linear basis B of M. This can be done by considering deterministic policies, namely those policies that prescribe a single action at each state with certainty. These are represented in the payoff space by pure state-action frequencies, as defined below.

Definition 4. A stationary policy $\boldsymbol{\pi}$ is deterministic if, for every state x, there exists an action a such that $\boldsymbol{\sigma}(a, x|\boldsymbol{\pi}) = 1$. A state-action frequency $\boldsymbol{\mu} \in M$ is *pure* if $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\pi}, x)$ for some deterministic policy $\boldsymbol{\pi}$ and initial state x. The set of pure state-action frequencies is denoted by M_0.

It can be shown that pure state-action frequencies are the extreme points of the convex set M.[14] Hence, every stationary state-action frequency can be obtained as a convex combination of pure ones. This implies that Condition 13 is equivalent to

$$M_0'\boldsymbol{u} = M_0'\nabla \boldsymbol{w}^*(\boldsymbol{\mu}) + k$$

and that there exists a linear basis B of M such that $B \subseteq M_0$. Hence a basis can be found selecting a maximal linearly independent subset of M_0, where M_0 can be computed by iterating all deterministic policies from every initial state.

[14]This was first proven by Derman [1970], and later by Hordijk and Kallenberg [1984] and Altman and Shwartz [1987]. It can be seen as an application to the current set-up of Kuhn's Theorem, stating the payoff equivalence of behavioral and mixed strategies in games with perfect recall [Kuhn, 1953].

3.1 Bregman divergence and Independence of Irrelevant Alternatives

Condition 13 characterizes the empirical content of undiscounted MDPs when a single instance of the decision problem is observed. It is also interesting to see which restrictions are imposed by the model on counterfactuals where specific alternatives are removed from agents' choice set at certain states. This Section defines a measure of statistical distance between state-action frequencies, and shows that Conditional Independence imposes a condition generalizing the well-known Independence of Irrelevant Alternatives [Luce, 1959], which requires that the relative distance of any pair of state-action frequencies from the observed one is independent of the alternatives available to agents. It then exploits this result to provide an axiomatic characterization of the undiscounted dynamic logit model.

To do this introduce, for every pair of stationary state-action frequencies $\boldsymbol{\mu}, \boldsymbol{\nu} \in M$, their *Bregman divergence* $D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu})$ associated with the generalized entropy $\boldsymbol{w}^*$. This is defined as the difference between the value of $\boldsymbol{w}^*$ at $\boldsymbol{\mu}$ and the value of the first-order Taylor expansion of $\boldsymbol{w}^*$ around $\boldsymbol{\nu}$ evaluated at $\boldsymbol{\mu}$:

$$D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu}) \equiv \boldsymbol{w}^*(\boldsymbol{\mu}) - \boldsymbol{w}^*(\boldsymbol{\nu}) - \nabla \boldsymbol{w}^*(\boldsymbol{\nu}) \cdot (\boldsymbol{\mu} - \boldsymbol{\nu}) \ .$$

Bregman divergences are often used as measures of statistical distance between probability distributions, as they satisfy a number of key properties.[15] Notice that we have $\boldsymbol{w}(\nabla \boldsymbol{w}^*(\boldsymbol{\nu})) = 0$ by Proposition 5. Hence, by Theorem 2, we have $\nabla \boldsymbol{w}^*(\boldsymbol{\nu}) \cdot \boldsymbol{\nu} = \boldsymbol{w}^*(\boldsymbol{\nu})$, and one can write

$$D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \boldsymbol{w}^*(\boldsymbol{\mu}) - \nabla \boldsymbol{w}^*(\boldsymbol{\nu}) \cdot \boldsymbol{\mu} = \sum_x \boldsymbol{\mu}_X(x)[w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) - \boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) \cdot \nabla w^*(\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x))].$$

In words, $D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is the $\boldsymbol{\mu}_X$-weighted average of the divergences between the choice probabilities associated with $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ at every state.

To state the restrictions imposed by Conditional Independence, one needs to define dynamic decision problems where specific alternatives are removed from agents' choice set A at certain states. To do this, define a decision sub-problem by means of a pair $\mathbb{D} = (\mathbb{A}, Y)$ such that $Y \subseteq X$ and $\mathbb{A} : \ Y \to 2^A$ is a

[15] For instance, $D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is always non-negative, and $D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$ if and only if $\boldsymbol{\mu} = \boldsymbol{\nu}$. Other properties include convexity in the first argument, a generalized Pythagorean theorem, and the fact that, given a random vector, the mean vector minimizes the expected Bregman divergence from the random vector (see Banerjee et al. 2005a). Bregman divergences were first introduced by Bregman [1967]. Their class include the squared Euclidean distance and the Kullback–Leibler divergence, among others. A summary of their properties can be found in Banerjee et al. [2005b].

correspondence mapping each state $x \in Y$ into the set of alternatives $\mathbb{A}(x) \subseteq A$ that are available to decision makers in that state. Denote by $\mathbb{D}^*$ the full decision problem where no alternative nor state is removed, that is, $\mathbb{D}^* = (\mathbb{A}^*, X)$ where $\mathbb{A}^*(x) = X$ for every x, and let $\mathcal{D}$ be a collection of decision problems such that $\mathbb{D}^* \in \mathcal{D}$ and such that each $(\mathbb{A}, Y) \in \mathcal{D}$ satisfies Accessibility. Namely, for every set of states $Y' \subsetneq Y$, there exist $y \in Y'$, $x \in Y \setminus Y'$ and $a \in \mathbb{A}(y)$ such that $T(x|a, y) > 0$.

For every $\mathbb{D} = (\mathbb{A}, Y) \in \mathcal{D}$, let $M^{\mathbb{D}} \subseteq M$ denote the set of stationary state-action frequencies of the decision problem associated with $\mathbb{D}$. That is, the set of all $\boldsymbol{\mu} \in M$ such that, for every state x, $\boldsymbol{\mu}(a, x) = 0$ if either $x \notin Y$ or $a \notin \mathbb{A}(x)$. Similarly, let also $M_0^{\mathbb{D}} \subseteq M_0$ be the set of pure state-action frequencies associated with $\mathbb{D}$. Suppose that, for every $\mathbb{D} \in \mathcal{D}$, the analyst observes a choice outcome $\boldsymbol{\mu}^{\mathbb{D}} \in M^{\mathbb{D}}$ resulting from agents' behavior when faced with decision problem $\mathbb{D}$, and let $\boldsymbol{\sigma}^{\mathbb{D}} \equiv \boldsymbol{\sigma}^{\boldsymbol{\mu}^{\mathbb{D}}}$ denote the associated system of conditional choice probabilities. This system of observations is consistent with the model if there exists some vector $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ of choice-specific payoffs such that, for every $\mathbb{D} \in \mathcal{D}$, $\boldsymbol{\mu}^{\mathbb{D}} = \boldsymbol{\mu}^{\mathbb{D}}(\boldsymbol{u})$ is the optimal state-action frequency for problem $\mathbb{D}$ under $\boldsymbol{u}$. By Proposition 5, this is the case if and only if

$$\forall\ \mathbb{D} \in \mathcal{D}\ \exists\ k \in \mathbb{R} :\ \boldsymbol{\mu} \cdot \boldsymbol{u} = \boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}}) + k\ \forall\ \boldsymbol{\mu} \in M_0^{\mathbb{D}}$$

In turn, it is easy to see that this implies the following:

$$\forall\ \mathbb{D}, \mathbb{D}' \in \mathcal{D},\ \forall\ \boldsymbol{\mu}, \boldsymbol{\nu} \in M_0^{\mathbb{D}} \cap M_0^{\mathbb{D}'} :\ D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\mu}^{\mathbb{D}}) - D_{\boldsymbol{w}^*}(\boldsymbol{\nu}, \boldsymbol{\mu}^{\mathbb{D}}) = D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\mu}^{\mathbb{D}'}) - D_{\boldsymbol{w}^*}(\boldsymbol{\nu}, \boldsymbol{\mu}^{\mathbb{D}'}) \tag{16}$$

It can be shows that the converse implication also holds. That is, Condition 16 fully characterizes the empirical content of the undiscounted MDP under a given distribution F of the i.i.d. shocks, when multiple decision problems are observed. I state this formally below.

Proposition 6. *Say that the system of observations $(\boldsymbol{\mu}^{\mathbb{D}})_{\mathbb{D} \in \mathcal{D}}$ - or, equivalently, the system of observations $(\boldsymbol{\sigma}^{\mathbb{D}})_{\mathbb{D} \in \mathcal{D}}$ - is rationalized by F if there exists $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ such that, for every $\mathbb{D} \in \mathcal{D}$, $\boldsymbol{\mu}^{\mathbb{D}}$ is rationalized by $\boldsymbol{u}$ in decision problem $\mathbb{D}$ when the i.i.d. shocks are distributed according to F. Then $(\boldsymbol{\mu}^{\mathbb{D}})_{\mathbb{D} \in \mathcal{D}}$ is rationalized by F if and only if Condition 16 is satisfied.*

Condition 16 can be seen as a generalization of Independence of Irrelevant Alternatives (IIA) to undiscounted MDPs, for a generic distribution F of the idiosyncratic shocks. In words, it requires that

the relative distance between any pair of pure state-action frequencies $\boldsymbol{\mu}, \boldsymbol{\nu}$ and the observed choice outcome $\boldsymbol{\mu}^{\mathbb{D}}$ is not affected by which other alternatives are available at a given decision problem. In other words, Conditional Independence imposes independence from irrelevant alternative state-action frequencies. It can be easily seen that, when F is the logit distribution, this naturally generalizes classic IIA to undiscounted MDPs. Indeed, in this case, $D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is the $\boldsymbol{\mu}_X$-weighted Kullback-Leibler divergence between $\boldsymbol{\sigma}^{\boldsymbol{\mu}}$ and $\boldsymbol{\sigma}^{\boldsymbol{\nu}}$ at each state:

$$D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_x \boldsymbol{\mu}_X(x) \sum_a \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a, x) \log \frac{\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a, x)}{\boldsymbol{\sigma}^{\boldsymbol{\nu}}(a, x)}$$

Hence Condition 16 can be written as

$$\forall\ \mathbb{D}, \mathbb{D}' \in \mathcal{D},\ \forall\ \boldsymbol{\mu}, \boldsymbol{\nu} \in M_0^{\mathbb{D}} \cap M_0^{\mathbb{D}'}:\ \sum_{a,x} [\boldsymbol{\mu}(a, x) - \boldsymbol{\nu}(a, x)] \log \boldsymbol{\sigma}^{\mathbb{D}}(a, x) = \sum_{a,x} [\boldsymbol{\mu}(a, x) - \boldsymbol{\nu}(a, x)] \log \boldsymbol{\sigma}^{\mathbb{D}'}(a, x) \tag{17}$$

Intuitively, for every $\boldsymbol{\mu} \in M$ and $\boldsymbol{\sigma} \in (\Delta A)^X$, the quantity $\sum_{a,x} \boldsymbol{\mu}(a, x) \log \boldsymbol{\sigma}(a, x)$ can be seen as the expected log-likelihood of $\boldsymbol{\sigma}$ when the data are drawn according to $\boldsymbol{\mu}$, a natural measure of how likely $\boldsymbol{\mu}$ is to be observed under $\boldsymbol{\sigma}$. Hence Condition 17 requires that the relative likelihood of observing $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ is not affected by which other alternatives are available. When there is a single state - i.e. $X = \{x\}$, so that every $\mathbb{D} = (\mathbb{A}^{\mathbb{D}}, \{x\}) \in \mathcal{D}$ is identified by a set $A^{\mathbb{D}} \equiv \mathbb{A}^{\mathbb{D}}(x)$ of available alternatives - this boils down to

$$\forall\ \mathbb{D}, \mathbb{D}' \in \mathcal{D},\ \forall\ a, a' \in A^{\mathbb{D}} \cap A^{\mathbb{D}'}:\ \frac{\boldsymbol{\sigma}^{\mathbb{D}}(a, x)}{\boldsymbol{\sigma}^{\mathbb{D}}(a', x)} = \frac{\boldsymbol{\sigma}^{\mathbb{D}'}(a, x)}{\boldsymbol{\sigma}^{\mathbb{D}'}(a', x)}$$

which is the usual condition for static models. From Proposition 6, Condition 17 fully characterizes the empirical content of the undiscounted dynamic logit model when multiple decision problems are observed. I state this formally below.

Corollary 1. *Say that the system of observations $(\boldsymbol{\sigma}^{\mathbb{D}})_{\mathbb{D} \in \mathcal{D}}$ is rationalized by the logit model if there exists $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ such that, for every $\mathbb{D} \in \mathcal{D}$, $\boldsymbol{\sigma}^{\mathbb{D}}$ is rationalized by $\boldsymbol{u}$ in decision problem $\mathbb{D}$ when F is the logit distribution. Then $(\boldsymbol{\sigma}^{\mathbb{D}})_{\mathbb{D} \in \mathcal{D}}$ is rationalized by the logit model if and only if Condition 17 is satisfied.*

In other words, Condition 17 axiomatically characterizes the undiscounted logit model, generalizing Luce's result to the dynamic set-up.

4 Identifying restrictions

Section 3 has shown that the vector $\boldsymbol{u}$ of choice-specific payoffs is not identified from aggregate choice data, unless additional restrictions are imposed. This Section characterizes the linear restrictions identifying $\boldsymbol{u}$, and discusses different types of restrictions commonly used in empirical work.

Before discussing identifying restrictions, it is useful to characterize the degree of under-identification of $\boldsymbol{u}$, namely the number of degrees of freedom of the identified set. From Section 3, recall that $\boldsymbol{u}$ rationalizes $\boldsymbol{\mu}$ if and only if $M_0'\boldsymbol{u} = M_0'\nabla \boldsymbol{w}^*(\boldsymbol{\mu}) + k$ for some constant $k \in \mathbb{R}$, where M_0 is the matrix with columns given by all pure state-action frequencies. Hence $\boldsymbol{\mu}$ identifies, up to a constant, an affine space of choice-specific payoff vectors of dimension equal to the dimension of the null space of M_0', which is equal to $|A|\,|X| - \dim M$. The following Lemma shows that this number is a function of the number of choices available at each state.

Lemma 1. *$dim M = (|A| - 1)\,|X| + 1$.*

In words, there exist $(|A|-1)\,|X|+1$ linearly independent state-action frequencies spanning the whole strategy space. In particular, recall from Section 3 that these can be chose to be pure, that is, associated with deterministic policies.

Example In Rust's engine replacement model, for every $1 \leq n \leq |X|$, let $\boldsymbol{\mu}^n$ be the state-action frequency associated with the deterministic policy that prescribes to replace the engine at x_m if and only if $n \leq m$, and let $\boldsymbol{\mu}^{|X|+1}$ be the state-action frequency associated with the deterministic policy that prescribes to never replace the engine. Letting $B = \{\boldsymbol{\mu}^n :\ n = 1, ..., |X| + 1\}$, it is easy to see that B is linearly independent, hence by Lemma 1 it spans the whole M.[16]

By Lemma 1, it follows that $\boldsymbol{\mu}$ identifies, up to a constant, an affine space of dimension $|X|-1$. Hence, intuitively, in order to identify $\boldsymbol{u}$ the analyst must impose $|X| - 1$ linear restrictions independent from M, and normalize one average payoff associated to some state-action frequency. In empirical practice,

[16] To see that B is linearly independent, pick $\boldsymbol{\mu}^n \in B$, and suppose by contradiction that $\boldsymbol{\mu}^n = \sum_{m \neq n} \alpha_m \boldsymbol{\mu}^m$ for some $\alpha \in \mathbb{R}^{|X|}$. Then we must have $\alpha_m = 0$ for every $m < n$. Indeed, if $n = 1$ this holds trivially. If $n > 1$, $\alpha_1 = 0$ since $\boldsymbol{\mu}^1(x_1, r) > 0$ and $\boldsymbol{\mu}^m(x_1, r) = 0$ for all $m > 1$. By induction, if $\alpha_m = 0$ for every $m < k < n$, then $\alpha_k = 0$, since $\boldsymbol{\mu}^k(x_k, r) > 0$ and $\boldsymbol{\mu}^m(x_k, r) = 0$ for all $m > k$. Similarly, we must have $\alpha_m = 0$ for every $m > n$. Indeed, if $n = |X| + 1$ this holds trivially. If $n < |X| + 1$, $\alpha_{|X|+1} = 0$ since $\boldsymbol{\mu}^{|X|+1}(x_{|X|}, nr) > 0$ and $\boldsymbol{\mu}^m(x_{|X|+1}, nr) = 0$ for all $m < |X| + 1$. By induction, if $\alpha_m = 0$ for every $m > k > n$, then $\alpha_k = 0$, since $\boldsymbol{\mu}^k(x_{k-1}, nr) > 0$ and $\boldsymbol{\mu}^m(x_{k-1}, nr) = 0$ for all $m < k$. Then we must have $\alpha = 0$, which implies $\boldsymbol{\mu}^n = 0_{|A||X|}$, a contradiction.

linear restrictions are typically imposed in the form of either exclusion or parametric restrictions. For some given matrix C with linearly independent columns in $\mathbb{R}^{|A||X|}$, exclusion restrictions usually take the form $C'\boldsymbol{u} = 0$, while parametric restrictions assume the there exists some $\theta \in \mathbb{R}^{|C|}$ to be estimated, such that $\boldsymbol{u} = C\theta$.

Definition 5. Given matrix C with linearly independent columns in $\mathbb{R}^{|A||X|}$, say that:

i) $C'\boldsymbol{u} = 0$ *identifies* $\boldsymbol{u}$ if, for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|A||X|}$, we have $\boldsymbol{u} = \boldsymbol{v}$ whenever $\boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{v})$ and $C'\boldsymbol{u} = C'\boldsymbol{v}$.

ii) $\boldsymbol{u} = C\theta$ identifies $\boldsymbol{u}$ if, for all $\theta, \delta \in \mathbb{R}^{|C|}$, we have $\theta = \delta$ whenever $\boldsymbol{\mu}(C\theta) = \boldsymbol{\mu}(C\delta)$.

Notice that, letting $C^{\perp}$ be an arbitrary basis of the null space of C, we have

$$C'\boldsymbol{u} = 0 \Leftrightarrow \exists\ \theta \in \mathbb{R}^{|C^{\perp}|} :\ \boldsymbol{u} = C^{\perp}\theta$$

hence exclusion and parametric restrictions are formally equivalent.[17] However, for completeness, I state conditions for identification for both types of restrictions.

Proposition 7. *1. The following are equivalent:*

(a) $C'\boldsymbol{u} = 0$ *identifies* $\boldsymbol{u}$

(b) The matrix $\begin{bmatrix} M_0' & 1_{|M_0|} \\ C' & 0_{|C|} \end{bmatrix}$ *has full column rank*

(c) C is such that: i) it includes $|X|-1$ *restrictions* $\{c^1, ..., c^{|X|-1}\} \subseteq C$ *such that* $Span\{c^1, ..., c^{|X|-1}\} \cap SpanM = \{0_{|A||X|}\}$*; and ii) there exists* $\boldsymbol{\nu} \in SpanC \cap SpanM$ *such that* $\sum_{a,x} \boldsymbol{\nu}(a,x) \neq 0$.

2. $\boldsymbol{u} = C\theta$ *identifies* $\boldsymbol{u}$ *if and only if the matrix* $\begin{bmatrix} M_0'C & \boldsymbol{1}_{|M_0|} \end{bmatrix}$ *has full column rank.*

For an arbitrary matrix C, Conditions 1.b and 2 can be easily verified numerically - in particular, the matrix M_0 can be computed by iterating all deterministic policies from every state. They follow directly from Proposition 5. To see this notice that, for exclusion restrictions, we have $\boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{v})$ and

[17] That is, $C^{\perp}$ is a matrix with $\left|C^{\perp}\right| = |A|\,|X| - |C|$ linearly independent columns in $\mathbb{R}^{|A||X|}$ such that $C'C^{\perp} = 0_{|C|\times|C^{\perp}|}$. Note that $C^{\perp}$ is a basis of the null space of C if and only if C is a basis of the nullspace of $C^{\perp}$, that is, we can write $(C^{\perp})^{\perp} = C$.

$C'\boldsymbol{u} = C'\boldsymbol{v}$ if and only if there exists $k \in \mathbb{R}$ such that $z = \begin{bmatrix} \boldsymbol{u} - \boldsymbol{v} \\ k \end{bmatrix}$ is a solution to the system

$$\begin{bmatrix} M_0' & \mathbf{1}_{|M_0|} \\ C' & \mathbf{0}_{|C|} \end{bmatrix} z = \mathbf{0}_{|M_0|+|C|}$$

Hence identification requires that the above system admits the unique solution $z = 0$, which is equivalent to Condition 1.b. Similarly, for parametric restrictions, we have $\boldsymbol{\mu}(C\theta) = \boldsymbol{\mu}(C\delta)$ if and only if there exists some $k \in \mathbb{R}$ such that $z = \begin{bmatrix} \theta - \delta \\ k \end{bmatrix}$ is a solution to the system

$$\begin{bmatrix} M_0'C & \mathbf{1}_{|M_0|} \end{bmatrix} z = \mathbf{0}_{|M_0|} \ .$$

Hence identification requires that the above system admits the unique solution $z = 0$, which is equivalent to Condition 2.

On the other hand, Condition 1.c formalizes the intuition behind identification. Namely, in order to identify $\boldsymbol{u}$, the analyst must impose: i) $|X| - 1$ linear restrictions which are independent from those already imposed by the data through M, and ii) normalize a linear combination of the average payoffs of stationary state-action frequencies. This normalization can either be imposed directly through C, or implied by combining the restrictions in C with those imposed by the data through M.

Finally, before moving to the examples, it is useful note that a minimal number of restrictions does not impose any constraint on the set of stationary measures that can be observed. Formally, the following holds.

Proposition 8. *Let C be such that $C'\boldsymbol{u} = 0$ (resp. $\boldsymbol{u} = C\theta$) identifies $\boldsymbol{u}$. If $|C| = |X|$ (resp. $|C| = (|A|-1)|X|$) then, for every $\boldsymbol{\mu} \in M_+$, there exists a unique $\boldsymbol{u}$ such that $C'\boldsymbol{u} = 0$ (resp. $\boldsymbol{u} = C'\theta$ for some $\theta \in \mathbb{R}^{|C|}$) and $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{u})$.*

If this is the case, I will say that $C'\boldsymbol{u} = 0$ (resp. $\boldsymbol{u} = C\theta$) *just-identifies* $\boldsymbol{u}$. Notice that this result holds for a generic distribution F of the idiosyncratic shocks satisfying Assumption 2. Hence, similarly to discounted models, aggregate choice data do not impose any restriction on F.

4.1 Examples

4.1.1 Normalizing one payoff at each state

One common type of restrictions is to impose that, for some action $a \in A$, we have $\boldsymbol{u}(a,x) = 0$ for all $x \in X$.[18] Intuitively, this amounts to assume that the payoff associated with action a is state-invariant - that is, $\boldsymbol{u}(a,x) = \boldsymbol{u}(a,x')$ for all x, x' - and to express all other coordinates of $\boldsymbol{u}$ in relative terms with respect to it. This is captured by the exclusion restrictions $C'\boldsymbol{u} = 0$, where

$$C = \{1\{a,x\} : \quad x \in X\}$$

Proposition 9. *$C'\boldsymbol{u} = 0$ identifies $\boldsymbol{u}$ if and only if one can find a state $x \in X$ such that, for every $Y \subseteq X \setminus \{x\}$, there exists $y \in Y$ such that $T(Y|a,y) < 1$.*

In words, identification requires the existence of a state x that can be reached in finite expected time from any other state $x' \neq x$ under the Markov state-transition probabilities $(T(a,x) : x \in X)$. Intuitively, this ensures that the set of restrictions $\{\boldsymbol{u}(a,x') = 0 \; \forall \; x' \neq x\}$ does not span any linear combination of stationary state-action frequencies - so that Condition 1.c.i of Proposition 7 is satisfied. Moreover, notice that C spans any stationary state-action frequency obtained by the deterministic policy prescribing action a at every state x, hence condition 1.c.ii of Proposition 7 is satisfied as well. It is interesting to note that Proposition 9 is in stark contrast with discounted models, for which it is known that normalizing one payoff at each state always identifies $\boldsymbol{u}$, regardless of the structure of transition probabilities. Hence it shows that, as far as identification is concerned, there is a discontinuity between discounted and undiscounted MDPs, and in particular that imposing a calibrated discount factor helps to identify agents' preferences.

Finally, note that $|C| = |X|$, hence, by Proposition 8, $C'\boldsymbol{u} = 0$ just-identifies $\boldsymbol{u}$. Moreover, it is easy to see that Proposition 9 does not rely on normalizing the same action at each state. That is, for any action profile $(a_x : x \in X)$, letting $C = \{1\{a_x, x\} : \quad x \in X\}$, the set of restrictions $C'\boldsymbol{u} = 0$ identifies u if and only if one can find a state $x \in X$ that can be reached with positive probability from any other state $x' \neq x$ under the transition probabilities $(T(a_x, x) : x \in X)$. Since by Accessibility one can

[18] The choice of normalizing the same action at each state is made simply for notational convenience, while everything in this Section would still hold if the action whose payoff is normalized changes depending on the state.

always find an action profile satisfying this property, Proposition 9 implies that there always exists a set of just-identifying restrictions.

4.1.2 Relative payoffs constant across states

Another common type of restrictions imposes that: i) for some pair of actions $a, a' \in A$, we have

$$\boldsymbol{u}(a', x) - \boldsymbol{u}(a, x) = \boldsymbol{u}(a', x') - \boldsymbol{u}(a, x') \text{ for every } x, x' \in X$$

and ii) some payoff is normalized to zero, that is, $\boldsymbol{u}(a'', x) = 0$ for some action $a'' \in A$ and state $x \in X$. Intuitively, this imposes that the relative payoff of taking action a' with respect to taking action a is state-independent, and all coordinates of $\boldsymbol{u}$ are express in relative terms with respect to $\boldsymbol{u}(a'', x)$. This is captured by the exclusion restrictions $C'\boldsymbol{u} = 0$, where

$$C = \{1\{a', x'\} - 1\{a, x'\} - (1\{a', x\} - 1\{a, x\}) : x' \neq x\} \cup \{1\{a'', x\}\}$$

Proposition 10. *Suppose that the set*

$$\{T(a', x') - T(a, x') - [T(a', x) - T(a, x)] : \quad x' \neq x\} \subseteq \mathbb{R}^X$$

is linearly independent. Then $C'\boldsymbol{u} = 0$ just-identifies $\boldsymbol{u}$.

Hence imposing that relative payoffs are constant across states plus a normalization identifies $\boldsymbol{u}$ except under knife-edge circumstances. Notice that, when $|A| = 2$, such as in the Rust's engine replacement model, these restrictions imply that payoffs are additive in actions and states. That is, there exists $\theta \in \mathbb{R}^{|A|+|X|}$ such that

$$\boldsymbol{u}(a, x) = \theta(a) + \theta(x) \text{ for all } a, x \ .$$

For a generic set A of actions, this is equivalent to impose that

$$\boldsymbol{u}(a', x) - \boldsymbol{u}(a, x) = \boldsymbol{u}(a', x') - \boldsymbol{u}(a, x') \quad \forall a, a', x, x' \ .$$

It follows that, in general, imposing additive payoffs - and normalizing the payoff of some action at some state x - identifies $\boldsymbol{u}$ if one can find a, a' such that the Condition of Proposition 10 holds.

Example Recall that, in Rust's engine replacement model, we imposed that $\boldsymbol{u}(a,x) = -c(x) - 1\{a = r\}RC$. It is easy to see that the set

$$\{T(r|x) - T(nr|x) - [T(r|x_1) - T(nr|x_1)] : \quad x \neq x_1\}$$

is linearly independent. Hence both RC and $c = (c(x))_{x \in X}$ are identified up to a normalization, such as $c(x_1) = 0$.

5 Estimation

The results of previous Sections suggest alternative two-steps strategies for estimating $\boldsymbol{u}$. As in Hotz and Miller [1993], the first step always consists in estimating the aggregate choice outcome in the population. The second step consists in inverting the estimated choice outcome, namely finding a vector of payoffs rationalizing it. Section 2 has shown that this is equivalent to solving a convex optimization. Section 5.1 will show that, under mild regularity conditions, a simple Tâtonnement procedure is guaranteed to converge monotonically to a solution of this problem. As shown in Section 3, however, this problem typically admits multiple solutions, hence additional restrictions must be imposed. If the imposed restrictions just-identify $\boldsymbol{u}$, then these restrictions always select a unique payoff vector rationalizing the estimated choice outcome. In the over-identified case - that is, when more restrictions are imposed than those strictly necessary to identify $\boldsymbol{u}$ - a payoff vector can be selected to minimize a measure of distance between model and data. This distance can either be in the space of payoffs or in the space of state-action frequencies. That is, $\boldsymbol{u}$ can be chosen either to minimize the distance between the payoff vectors satisfying the imposed restrictions and those rationalizing the data, or to minimize a measure of distance between the observed state-action frequency and that rationalized by the model. These two cases are treated in Sections 5.2.1 and 5.2.2, which consider an estimator minimizing the Euclidean distance in the space of payoffs, and an estimator minimizing the Bregman divergence in the space of state-action frequencies, respectively.

5.1 Inversion

In this Section, fix an arbitrary state-action frequency $\boldsymbol{\mu} \in M_+$ and the associated system $\boldsymbol{\sigma} = \boldsymbol{\sigma}^{\boldsymbol{\mu}}$ of conditional choice probabilities. The inversion problem consists in finding a vector $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ such that $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{u})$ or, equivalently, $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{u})$. As shown in Section 2, this is equivalent to finding a solution to Problem 11, which admits multiple solutions. As shown by Proposition 5, a particular solution is given by the gradient $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ of the generalized entropy at $\boldsymbol{\mu}$. In alternative, a particular solution can be selected by imposing additional linear restrictions on $\boldsymbol{u}$. This Section shows convergence of simple algorithms performing these two types of inversion. The results rely on a mild additional assumptions on the distribution F of the idiosyncratic shocks. Recall that, by Assumption 2, for every pair of actions $a, a' \in A$, the distribution of the random variable $\epsilon_{a'} - \epsilon_a$ admits a density with full-support on the real line. I will require such density to be bounded above.

Definition 6. Say that F is *regular* if, for every $a, a' \in A$ with $a \neq a'$, the distribution of $\epsilon_{a'} - \epsilon_a$ admits a density bounded above.

5.1.1 Computing the gradient of $\boldsymbol{w}^*$

When F is chosen so that $\boldsymbol{w}^*$ is not known in closed form, the gradient $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ can be easily computed by means of standard convex optimization methods. This is a consequence of the following result.

Lemma 2. $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ *is the unique vector* $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ *satisfying*

$$\boldsymbol{u}(x) \in \arg\max_{u \in \mathbb{R}^A} [\boldsymbol{\sigma}(x) \cdot u - w(u)] \text{ and } w(\boldsymbol{u}(x)) = 0 \text{ for all } x \in X \tag{18}$$

In other words, by Theorem 1, $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ is the unique vector $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ such that, for every state x, $\boldsymbol{u}(x)$ rationalizes the vector $\boldsymbol{\sigma}(x)$ of choice probabilities at x in the static sense - that is, as in Section 2.1 - and $w(\boldsymbol{u}(x)) = 0$. Hence computing $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ boils down to solve one convex optimization problem for each state. Fix a state x, and consider the iterates $(u^n)_{n \geq 0}$ generated by the gradient method with constant step size $\gamma > 0$ applied to Problem 18, normalized to that $w(u^n) = 0$ at every step. From an

arbitrary starting point $u^0 \in \mathbb{R}^A$, these can be written as

$$u^{n+1} = u^n + \gamma[\boldsymbol{\sigma}(x) - \sigma^n] - w(u^n + \gamma[\boldsymbol{\sigma}(x) - \sigma^n]) \text{ for every } n \geq 0 \tag{19}$$

where, for every $n \geq 0$, $\sigma^n \in \Delta A$ denotes the vector of choice probabilities rationalized by u^n:

$$\sigma^n(a) = \mathrm{P}_F[u^n(a) + \epsilon(a) = \max_{a'}[u^n(a') + \epsilon(a')]] \text{ for every } a \in A \ .$$

Indeed, recall that Theorem 1 ensures that $\nabla w(u^n) = \sigma^n$ for every $n \geq 0$.

Proposition 11. *Suppose that F is regular. Then there exists $\bar{\gamma} > 0$ such that, for all $0 < \gamma \leq \bar{\gamma}$, we have*

$$\boldsymbol{\sigma}(x) \cdot u^{n+1} \geq \boldsymbol{\sigma}(x) \cdot u^n + \frac{\gamma}{2} \left\| \boldsymbol{\sigma}(x) - \sigma^n \right\|_2^2 \ .$$

In words, under regularity, the iterates in 19 converge monotonically to $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})(x)$. The proof consists in showing that if F is regular then w is a smooth function, hence the progress bounds can be derived using standard arguments.[19]

5.1.2 Solving the dual under linear restrictions

This Section shows how to compute a solution of the dual Problem 11 under a given set of linear restrictions identifying $\boldsymbol{u}$. For notational convenience, I focus on the case in which the restrictions are written in parametric form: $\boldsymbol{u} = C\theta$ for some arbitrary matrix C such that $\boldsymbol{u} = C\theta$ identifies $\boldsymbol{u}$. The goal is to compute the unique solution θ^* of

$$\max_{\theta \in \mathbb{R}^{|C|}} [\boldsymbol{\mu} \cdot C\theta - \boldsymbol{w}(C\theta)] \ . \tag{20}$$

That is, $\boldsymbol{u}^* = C\theta^*$ solves the dual Problem 11 under the constraint that $\boldsymbol{u} = C\theta$ for some $\theta \in \mathbb{R}^{|C|}$. Theorem 2 then ensures that, if there exists θ such that $\boldsymbol{\mu}(C\theta) = \boldsymbol{\mu}$, then $\theta^* = \theta$. In particular, if $\boldsymbol{u} = C\theta$ just-identifies $\boldsymbol{u}$ - that is, when $|C| = (|A| - 1)\,|X|$ - θ^* always equals the unique vector of parameters rationalizing $\boldsymbol{\mu}$. In general, Problem 20 is strictly convex and differentiable, and its solution

[19] See Appendix A for a definition of smooth functions.

is characterized by the first-order conditions

$$C'[\boldsymbol{\mu} - \boldsymbol{\mu}(C\theta^*)] = 0 \ .$$

Indeed, Theorem 2 ensures that, for every $\theta \in \mathbb{R}^{|C|}$, the gradient of $\boldsymbol{w}(C\theta)$ with respect to θ equals to $C'\boldsymbol{\mu}(C\theta)$. Consider the iterates $(\theta^n)_{n\geq 0}$ generated by the gradient method with constant step size $\gamma > 0$ applied to Problem 20. From an arbitrary starting point $\theta^0 \in \mathbb{R}^{|C|}$, these can be written as

$$\theta^{n+1} = \theta^n + \gamma C'[\boldsymbol{\mu} - \boldsymbol{\mu}(C\theta^n)] \text{ for every } n \geq 0.$$

The following result states that, under regularity, these iterates converge monotonically to θ^*.

Proposition 12. *Suppose that F is regular. Then there exists $\bar{\gamma} > 0$ such that, for all $0 < \gamma \leq \bar{\gamma}$, we have*

$$\boldsymbol{\mu} \cdot C\theta^{n+1} - \boldsymbol{w}(C\theta^{n+1}) \geq \boldsymbol{\mu} \cdot C\theta^n - \boldsymbol{w}(C\theta^n) + \frac{\gamma}{2} \left\| C'[\boldsymbol{\mu} - \boldsymbol{\mu}(C\theta^n)] \right\|_2^2 .$$

The intuition behind this result is slightly more complex than the one of Proposition 11, due to the fact that regularity of F does not guarantee smoothness of the function $\boldsymbol{w}$. Instead the proof shows that, provided that γ is chosen so that all coordinates of $\boldsymbol{\mu}_X(C\theta^n)$ remain bounded away from zero, the variations $\left\|\boldsymbol{\mu}(C\theta^{n+1}) - \boldsymbol{\mu}(C\theta^n)\right\|_2$ in state-action frequencies can be bounded by a factor of the associated variations $\left\|\boldsymbol{\sigma}(C\theta^{n+1}) - \boldsymbol{\sigma}(C\theta^n)\right\|_2$ in the associated conditional choice probabilities. Combined with the smoothness of w, this yields the progress bounds for γ small enough.

5.2 Estimation

For the remainder of this Section, fix a set of linear restrictions identifying $\boldsymbol{u}$. For notational convenience, I focus on the case in which the restrictions are written in parametric form, so that $\boldsymbol{u} = C\theta$ for some arbitrary matrix C such that $\boldsymbol{u} = C\theta$ identifies $\boldsymbol{u}$. The results so far suggest two alternative strategies for estimating θ. In both cases, as in Hotz and Miller [1993], the first step consists in estimating the aggregate choice outcomes in the population. In what follows, I denote by $\hat{\boldsymbol{\sigma}} \in (\Delta A)^X$ an estimate of the conditional choice probabilities, and by $\hat{\boldsymbol{\mu}}$ an estimate of the associated stationary state-action frequencies.

In particular, I assume that both $\hat{\boldsymbol{\sigma}}$ and $\hat{\boldsymbol{\mu}}$ have full support.

5.2.1 Constrained Least Squares

Consider an estimator minimizing the distance between the set of choice-specific payoff vectors satisfying the imposed restrictions and those rationalizing $\hat{\boldsymbol{\mu}}$. Formally, let $\|\cdot\|$ be some norm on $\mathbb{R}^{|A||X|}$, and consider an estimator solving the Constrained Least Squares problem

$$\min_{\theta\in\mathbb{R}^{|C|},\boldsymbol{v}\in\mathbb{R}^{|A||X|}} \|\boldsymbol{v}-C\theta\|^2 \text{ s.t. } \boldsymbol{\mu}(\boldsymbol{v})=\hat{\boldsymbol{\mu}} \ . \tag{21}$$

Let B be a linear basis for M, which, as shown in Sections 3 and 4, can be taken to be a set of $(|A|-1)\,|X|+1$ linearly independent pure state-action frequencies. Then the constraints of Problem 21 can be written linearly in $\boldsymbol{v}$, so that this is equivalent to

$$\min_{\theta\in\mathbb{R}^{|C|},\boldsymbol{v}\in\mathbb{R}^{|A||X|},k\in\mathbb{R}} \|\boldsymbol{v}-C\theta\|^2 \text{ s.t. } B'\boldsymbol{v}=B'\nabla\boldsymbol{w}^*(\hat{\boldsymbol{\mu}})+k \ . \tag{22}$$

Note that, since $\boldsymbol{u}=C\theta$ identifies $\boldsymbol{u}$, this problem admits a unique solution $\hat{\theta},\hat{\boldsymbol{v}},\hat{k}$, which, by Proposition 5, is such that $\hat{k}=\boldsymbol{w}(\hat{\boldsymbol{v}})$.[20] In particular, if $\boldsymbol{u}=C\theta$ just-identifies $\boldsymbol{u}$ - that is, if $|C|=(|A|-1)\,|X|$ - this is equivalent to solve for the unique $\hat{\theta}$ such that $C\hat{\theta}$ rationalizes $\hat{\boldsymbol{\mu}}$: letting $D\equiv\begin{bmatrix} B'C & 1_{|M_0|}\end{bmatrix}$, D is invertible by Proposition 7, hence we can write

$$\begin{bmatrix}\hat{\theta}\\ -\hat{k}\end{bmatrix}=D^{-1}B'\nabla\boldsymbol{w}^*(\hat{\boldsymbol{\mu}}) \text{ and } \hat{\boldsymbol{v}}=C\hat{\theta}$$

[20] To see that the solution is unique, suppose that $\theta^1,\boldsymbol{v}^1,k^1$ and $\theta^2,\boldsymbol{v}^2,k^2$ are two solutions. Then we must have $\boldsymbol{v}^1-C\theta^1=\boldsymbol{v}^2-C\theta^2$, since otherwise, by strict convexity of the squared norm, this would yield the contradiction

$$\left\|\frac{\boldsymbol{v}^1+\boldsymbol{v}^2}{2}-C\frac{\theta^1+\theta^2}{2}\right\|^2<\frac{\left\|\boldsymbol{v}^1-C\theta^1\right\|^2+\left\|\boldsymbol{v}^2-C\theta^2\right\|^2}{2}.$$

This implies that $B'C\theta^1=B'C\theta^2+k^1-k^2$, hence $\boldsymbol{\mu}(C\theta^1)=\boldsymbol{\mu}(C\theta^2)$. Since $\boldsymbol{u}=C\theta$ identifies $\boldsymbol{u}$, this implies $\theta^1=\theta^2$, hence $\boldsymbol{v}^1=\boldsymbol{v}^2$ and $k^1=k^2$.

In the over-identified case - that is, when $|C| < (|A| - 1)|X|$ - the solution can still be written in closed form whenever $\|\cdot\|$ is an inner product norm.[21] For instance, when $\|\cdot\| = \|\cdot\|_2$ we have

$$\begin{bmatrix} \hat{\theta} \\ -\hat{k} \end{bmatrix} = [D'[B'B]^{-1}D]^{-1}D'[B'B]^{-1}B'\nabla \boldsymbol{w}^*(\hat{\boldsymbol{\mu}})$$

and

$$\hat{\boldsymbol{v}} = C\hat{\theta} + B[B'B]^{-1}B'[\nabla \boldsymbol{w}^*(\hat{\boldsymbol{\mu}}) + \hat{k} - C\hat{\theta}]$$

In conclusion, θ can be estimated though the following steps: i) compute the set M_0 of stationary state-action frequencies associated with pure policies, and select a maximal linearly independent subset $B \subseteq M_0$, ii) compute $\nabla \boldsymbol{w}^*(\hat{\boldsymbol{\mu}})$ as in Section 5.1.1, and iii) solve for $\hat{\theta}$ as above. This procedure yields an estimator minimizing the distance between model and data in the space of payoffs.

5.2.2 Bregman projection

The Constrained Least Squares estimator discussed in previous Section was devised to minimize the distance between model and data in the space of payoffs. This Section considers instead a measure of their distance in the space of outcomes. Section 3.1 introduced the Bregman divergence $D_{w^*}(\boldsymbol{\mu}, \boldsymbol{\nu})$ associated with the generalized entropy $\boldsymbol{w}^*$ between two arbitrary stationary state-action frequencies $\boldsymbol{\mu}, \boldsymbol{\nu} \in M$, showing that this can be written as

$$D_{\boldsymbol{w}^*}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \boldsymbol{w}^*(\boldsymbol{\mu}) - \boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\nu}) \ .$$

In the context of estimation, the divergence $D_{\boldsymbol{w}^*}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}(\boldsymbol{u}))$ between the estimated state-action frequency and that rationalized by a given vector $\boldsymbol{u}$ of payoffs provides a natural objective. Indeed, by Proposition 5, we have $\hat{\boldsymbol{\mu}} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u})) = \hat{\boldsymbol{\mu}} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u})$, so that minimizing $D_{\boldsymbol{w}^*}(\hat{\boldsymbol{\mu}}, \boldsymbol{\mu}(\boldsymbol{u}))$ with respect to $\boldsymbol{u}$ is equivalent to maximizing $\hat{\boldsymbol{\mu}} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u})$. Under the parametric restrictions $\boldsymbol{u} = C\theta$, this yields the estimator

$$\hat{\theta} = \arg \max_{\theta \in \mathbb{R}^{|C|}} [\hat{\boldsymbol{\mu}} \cdot C\theta - \boldsymbol{w}(C\theta)] \ .$$

[21] See for instance Amemiya [1985, Section 1.4.1].

In other words, $\hat{\theta}$ minimizes the Bregman divergence between the estimated state-action frequency $\hat{\boldsymbol{\mu}}$ and those rationalized by some vector of parameters θ. It is interesting to note that, when F is a standard logit distribution and the data are assumed to be sampled from a stationary distribution, $\hat{\theta}$ is the maximum likelihood estimator. Formally, suppose that the data consist of N observations a_i, x_i, $i = 1, ..., N$, where each observation i describes the action a_i taken by an agent in state x_i. If these observations are drawn from their stationary distribution, a consistent estimate for the state-action frequency is given by

$$\hat{\boldsymbol{\mu}}(a, x) = \sum_{i=1}^{N} \frac{1\{a_i = a, x_i = x\}}{N} \text{ for every } a, x.$$

therefore, if F is a standard Logit distribution, we have

$$\hat{\boldsymbol{\mu}} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}(C\theta)) = \frac{1}{N} \sum_{i=1}^{N} 1\{a_i = a, x_i = x\} \log \boldsymbol{\sigma}(a_i, x_i | C\theta)$$

so that $\hat{\theta}$ is the maximum likelihood estimator.

In conclusion, the dual structure of undiscounted MDPs suggests alternative estimation strategies, depending on whether the analyst wishes to target the distance between model and data in the space of payoffs or in the space of outcomes. Both procedures ultimately boil down to solving well-behaved convex optimization problems. In the just-identified case, these procedures are equivalent, yielding alternative routines for computing inverting the observed choice outcomes.

6 Extensions

The dual framework makes it especially convenient to work with state-action frequencies as opposed to conditional choice probabilities. In particular, many of the results presented so far extend directly to situations where the the analyst's observations are obtained by aggregating state-action frequencies in a linear fashion. This Section exemplifies this by presenting two instances of such linear aggregation. First, I consider the case in which the analyst observes a state-action frequency obtained by averaging state-action frequencies of agents of different types. The type of each agent is represented by a vector $\boldsymbol{u}$ of choice-specific payoffs, where $\boldsymbol{u}$ is distributed in the population according to some distribution G, and the analyst observes $\int \boldsymbol{\mu}(\boldsymbol{u}) \mathrm{d}G$. In this case, the results on identification apply to the average vector $\int \boldsymbol{u} \mathrm{d}G$

of choice-specific payoffs in the population. Second, I consider the case in which the analyst cannot distinguish every action and state visited by agents, but instead observes the cumulative frequencies $\sum_{(a,x)\in z} \boldsymbol{\mu}(a,x)$ of a coarser partition Z of $A \times X$. In this case, the results on identification and inversion apply to the average payoffs $(1/|z|) \cdot \sum_{(a,x)\in z} \boldsymbol{u}(a,x)$ attached to each cell of the partition. This Section keeps an informal stlyle since, except for Propositions 13 and 14, the results follow from the exact same reasoning as in the baseline model. Hence proofs are provided in the Appendix for Propositions 13 and 14 only.

6.1 Mixed models

This Section considers the case in which agents' preferences are affected by correlated fixed effects. Let the type of each agent be described by a vector $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ of choice-specific payoffs, constant over time. Each agent observes his type, which is not observed by the econometrician. Let $\bar{\boldsymbol{u}}$ denote the average type in the population, and $\boldsymbol{\eta} \equiv \boldsymbol{u} - \bar{\boldsymbol{u}}$ be distributed in the population according to some distribution G. Then we can write

$$\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{\eta} \text{ where } \boldsymbol{\eta} \sim G \text{ with } \mathrm{E}_G \boldsymbol{\eta} = 0.$$

In words, fixed effects are now the sum of an idiosyncratic component ϵ, drawn each period according to F, and a persistent component $\boldsymbol{\eta}$, distributed in the population according to G. Given this, the state space is modified and described by $\boldsymbol{\eta}, x, \epsilon$. The per-period utility that an agent of type $\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{\eta}$ derives from choosing a in state x is

$$\bar{\boldsymbol{u}}(a,x) + \boldsymbol{\eta}(a,x) + \epsilon(a) \ .$$

Conditional on the current state x and choice a, the future value x' of x is drawn with probability $T(x'|a,x)$. The future value ϵ' of ϵ is then drawn from F, and the next state of the agent is $\boldsymbol{\eta}, x', \epsilon'$. In what follows, I fix two distributions F and G, and consider the problem of estimating $\bar{\boldsymbol{u}}$ from the aggregate choice outcomes, treating F and G as known. Typically, in discounted models, difficulties with respect to the homogeneous case arise from the present of dynamic selection, since the state variable $\boldsymbol{\eta}$ is unobserved by the analyst, hence the observed histories of states and choices are no longer independent from the unobservables.[22] However, this Section shows that for undiscounted MDPs many of the results

[22] See for instance Cameron and Heckman [1998], Taber [2000].

presented so far generalize easily, since state action frequencies aggregate linearly. Formally, say that $\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}$ *rationalizes* $\boldsymbol{\mu} \in M$, written $\boldsymbol{\mu} = \boldsymbol{\mu}(\bar{\boldsymbol{u}}; G)$, if it holds

$$\boldsymbol{\mu} = \int \boldsymbol{\mu}(\bar{\boldsymbol{u}} + \boldsymbol{\eta}) \mathrm{d}G(\boldsymbol{\eta}).$$

In words, for every a, x and every $\boldsymbol{\eta}$ in the support of G, $\boldsymbol{\mu}(a, x|\bar{\boldsymbol{u}} + \boldsymbol{\eta})$ is the frequency according to which an agent visits state x and takes action a in the long-run, conditional on the agent's type being $\bar{\boldsymbol{u}} + \boldsymbol{\eta}$. The analyst does not observe agents' types, but only the unconditional state-action frequencies $\int \boldsymbol{\mu}(a, x|\bar{\boldsymbol{u}} + \boldsymbol{\eta}) \mathrm{d}G(\boldsymbol{\eta})$ obtained by integrating the conditional frequencies across the distribution of types in the population. By Theorem 2 it follows that $\bar{\boldsymbol{u}}$ rationalizes $\boldsymbol{\mu}$ if and only if $\boldsymbol{\mu} = \int \nabla \boldsymbol{w}(\bar{\boldsymbol{u}} + \boldsymbol{\eta}) \mathrm{d}G(\boldsymbol{\eta})$. Since $\boldsymbol{w}(\cdot)$ and $\nabla \boldsymbol{w}(\cdot)$ are both continuous functions, this is equivalent to

$$\boldsymbol{\mu} = \nabla \boldsymbol{w}(\bar{\boldsymbol{u}}; G)$$

where

$$\boldsymbol{w}(\bar{\boldsymbol{u}}; G) \equiv \int \boldsymbol{w}(\bar{\boldsymbol{u}} + \boldsymbol{\eta}) \mathrm{d}G(\boldsymbol{\eta})$$

is the average value of the undiscounted MDP in the population. These are the first order conditions of the Problem

$$\max_{\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}} \left[\boldsymbol{\mu} \cdot \bar{\boldsymbol{u}} - \boldsymbol{w}(\bar{\boldsymbol{u}}; G) \right] . \tag{23}$$

Hence, in analogy with the case with uncorrelated fixed effects, the set of average payoff vectors rationalizing a given state-action frequency is the solution set of a convex optimization.

6.1.1 Identification

The dual Problem 23 can be exploited to show that the results on identification extend directly to models with persistent heterogeneity.

Proposition 13. *For any $\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}} \in \mathbb{R}^{|A||X|}$, we have $\boldsymbol{\mu}(\bar{\boldsymbol{u}}; G) = \boldsymbol{\mu}(\bar{\boldsymbol{v}}; G)$ if and only if there exists a scalar*

$k \in \mathbb{R}$ *such that*

$$\boldsymbol{\nu} \cdot \bar{\boldsymbol{u}} = \boldsymbol{\nu} \cdot \bar{\boldsymbol{v}} + k \text{ for every } \boldsymbol{\nu} \in M. \tag{24}$$

In words, similarly to homogeneous models, $\boldsymbol{\mu}$ identifies the average payoff associated to every stationary state-action frequency. The proof exploits Proposition 5 to show that, for every $\bar{\boldsymbol{u}} \neq \bar{\boldsymbol{v}}$, $\boldsymbol{w}(\cdot; G)$ is strictly convex on the segment $[\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}]$ whenever Condition 24 fails, hence Problem 23 cannot admit both $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{v}}$ as distinct solutions. Conversely, by Proposition 5, Condition 24 ensures that $\boldsymbol{\mu}(\bar{\boldsymbol{u}}+\boldsymbol{\eta}) = \boldsymbol{\mu}(\bar{\boldsymbol{v}}+\boldsymbol{\eta})$, for every $\boldsymbol{\eta}$ in the support of G, so that $\boldsymbol{\mu}(\bar{\boldsymbol{u}}; G) = \boldsymbol{\mu}(\bar{\boldsymbol{v}}; G)$.

A consequence of this is that the results on the identifying power of linear restrictions extend directly as well. To state them formally, similarly to Section 4, if C is a matrix with columns in $\mathbb{R}^{|A||X|}$, say that $C'\bar{\boldsymbol{u}} = 0$ identifies $\bar{\boldsymbol{u}}$ if, for all $\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}} \in \mathbb{R}^{|A||X|}$, we have $\bar{\boldsymbol{u}} = \bar{\boldsymbol{v}}$ whenever $C'\bar{\boldsymbol{u}} = C'\bar{\boldsymbol{v}}$ and $\boldsymbol{\mu}(\bar{\boldsymbol{u}}; G) = \boldsymbol{\mu}(\bar{\boldsymbol{v}}; G)$, and that $\bar{\boldsymbol{u}} = C\theta$ identifies $\bar{\boldsymbol{u}}$ if, for all $\theta, \delta \in \mathbb{R}^{|C|}$, we have $\theta = \delta$ whenever $\boldsymbol{\mu}(C\theta; G) = \boldsymbol{\mu}(C\delta; G)$. Then, replacing $\boldsymbol{u}$ with $\bar{\boldsymbol{u}}$ in the statement of Proposition 7, the result still holds. Given this, the under-identification result of Proposition 8 generalizes as well. Formally, the following holds.

Proposition 14. *If C is such that $C'\bar{\boldsymbol{u}} = 0$ (resp. $\bar{\boldsymbol{u}} = C\theta$) identifies $\bar{\boldsymbol{u}}$ and $|C| = |X|$ (resp. $|C| = (|A|-1)|X|$) then, for every $\boldsymbol{\mu} \in M_+$, there exists a unique $\bar{\boldsymbol{u}}$ such that $C'\bar{\boldsymbol{u}} = 0$ (resp. $\bar{\boldsymbol{u}} = C'\theta$ for some $\theta \in \mathbb{R}^{|C|}$) and $\boldsymbol{\mu} = \boldsymbol{\mu}(\bar{\boldsymbol{u}}; G)$.*

Interestingly, notice that non of this requires making any assumption on the distribution G.

6.1.2 Inversion

Another consequence of the dual Problem 23 is that the inversion procedure described in Section 5.1.2 generalizes directly to mixed models. Formally, if C is such that $\bar{\boldsymbol{u}} = C\theta$ identifies $\bar{\boldsymbol{u}}$ then, for any given $\boldsymbol{\mu} \in M_+$, the problem

$$\max_{\theta \in \mathbb{R}^{|C|}} [\boldsymbol{\mu} \cdot C\theta - \boldsymbol{w}(C\theta; G)]$$

admits a unique solution θ^*. This can computed for instance through the gradient method, which, for a given constant $\gamma > 0$ and an arbitrary starting point $\theta^0 \in \mathbb{R}^{|C|}$, generates the iterates

$$\theta^{n+1} = \theta^n + \gamma C'[\boldsymbol{\mu} - \boldsymbol{\mu}(C\theta^n; G)] \text{ for every } n \geq 0.$$

If $|C| = (|A| - 1)|X|$ then θ^* is the unique vector of parameters rationalizing $\boldsymbol{\mu}$, that is, such that $\boldsymbol{\mu} = \boldsymbol{\mu}(C\theta^*; G)$. Otherwise, defining $\boldsymbol{w}^*(\boldsymbol{\mu}; G)$ as the value of Problem 23, θ^* minimizes the Bregman divergence $D_{\boldsymbol{w}^*(\cdot;G)}(\boldsymbol{\mu}, \boldsymbol{\mu}(C\theta; G))$ between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}(C\theta; G)$. Hence, in principle, this procedure can be used to compute a vector $\bar{\boldsymbol{u}}^*$ rationalizing $\boldsymbol{\mu}$, so that θ can be estimated via constrained least squares, replacing $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ with $\bar{\boldsymbol{u}}^*$ in the treatment of Section 5.2.1, or to compute a Bregman projection estimator as in Section 5.2.2.

6.2 Linear feedback

Let Z be a partition of $A \times X$, and consider the case in which the analyst cannot distinguish every action and state visited by agents, but instead observes the cumulative frequencies $\bar{\boldsymbol{\mu}} \in \mathbb{R}^Z$, where $\bar{\boldsymbol{\mu}}(z) = \sum_{(a,x)\in z} \boldsymbol{\mu}(a, x)$ of every z. Let $\bar{\boldsymbol{u}} \in \mathbb{R}^{|Z|}$ be defined by $\bar{\boldsymbol{u}}(z) = (1/|z|) \cdot \sum_{(a,x)\in z} \boldsymbol{u}(a, x)$ for every z. That is, $\bar{\boldsymbol{u}}(z)$ denotes the average payoff attached to z. Let also $\boldsymbol{\eta} \in \mathbb{R}^{|A||X|}$ be defined by

$$\boldsymbol{\eta}(a, x) = \boldsymbol{u}(a, x) - |z|\, \bar{\boldsymbol{u}}(z) \text{ for every } z \in Z \text{ and every } (a, x) \in z$$

and $E \in \{0, 1\}^{|A||X|\times|Z|}$ be the matrix such that, for every $(a, x), z$, $E_{(a,x),z} = 1$ if $(a, x) \in z$ and $E_{(a,x),z} = 0$ otherwise. Then one can write

$$\bar{\boldsymbol{\mu}} = E'\boldsymbol{\mu} \text{ and } \boldsymbol{u} = E\bar{\boldsymbol{u}} + \boldsymbol{\eta} \ .$$

Intuitively, $\bar{\boldsymbol{\mu}}$ conveys information about the full state-action frequency $\boldsymbol{\mu}$. For this reason, in what follow I will refer to $\bar{\boldsymbol{\mu}}$ as a *feedback*, and denote by

$$\bar{M} \equiv \{E'\boldsymbol{\mu} : \ \boldsymbol{\mu} \in M\}$$

the set of all observable feedbacks. In turn, $\boldsymbol{\mu}$ conveys information about the choice-specific payoffs. Hence $\bar{\boldsymbol{\mu}}$ conveys information about $\boldsymbol{u}$ as well, although some of the information contained in $\boldsymbol{\mu}$ is lost in the aggregation. This Section shows that, intuitively, the information conveyed by $\bar{\boldsymbol{\mu}}$ is encapsulated by $\bar{\boldsymbol{u}}$. That is, the previous results extend directly to the case in which the analyst makes inference on $\bar{\boldsymbol{u}}$ from the observation of $\bar{\boldsymbol{\mu}}$.

Example 1 Consider an analyst that can observe agents' states, but not their actions. This is captured by $Z = X$, and $E_{(a,x),y} = 1$ if $y = x$ and 0 otherwise, so that $\bar{\boldsymbol{\mu}}(x) = \boldsymbol{\mu}_X(x) = \sum_a \boldsymbol{\mu}(a,x)$, and $\bar{\boldsymbol{u}}(x) = (1/|z|)\sum_a \boldsymbol{u}(a,x)$. $\bar{M}$ is the set of all stationary distributions over states associated with some policy.

Example 2 Consider an analyst that can only observe agents' state conditionally on them taking a particular action. For instance, consider the engine replacement model, and suppose that the analyst can only observe the mileage of buses replacing the engine. If the analyst also knows the total fleet size, this is equivalent to observing the state-action frequencies $\boldsymbol{\mu}(r,x)$ for all states x conditional on replacing the engine, and the cumulative frequency $\sum_x \boldsymbol{\mu}(nr,x)$ according to which the engine is not replaced. This situation is captured by $Z = \{(r,x) : x \in X\} \cup \{nr\}$, $E_{(r,x),z} = 1$ if $z = (r,x)$ and 0 otherwise, and $E_{(nr,x),z} = 1$ if $z = nr$ and 0 otherwise. The average payoffs are given by $\bar{\boldsymbol{u}}(r,x) = -c(x) - RC$ for all x and $\bar{\boldsymbol{u}}(nr) = -(1/|X|)\sum_x c(x)$.

Consider the situation in which $\boldsymbol{\eta}$ is known, and the analyst makes inference on $\bar{\boldsymbol{u}}$ based on the observation of $\bar{\boldsymbol{\mu}}$. Say that $\bar{\boldsymbol{u}}$ rationalizes $\bar{\boldsymbol{\mu}}$, written $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}(\bar{\boldsymbol{u}})$, if $\bar{\boldsymbol{\mu}} = E'\boldsymbol{\mu}(E\bar{\boldsymbol{u}} + \boldsymbol{\eta})$. By Theorem 2, this is equivalent to $\bar{\boldsymbol{\mu}}$ solving

$$\max_{\bar{\boldsymbol{\mu}} \in \bar{M}} \{\bar{\boldsymbol{\mu}} \cdot \bar{\boldsymbol{u}} - \{\min_{\boldsymbol{\mu} \in M} [\boldsymbol{w}^*(\boldsymbol{\mu}) - \boldsymbol{\mu} \cdot \boldsymbol{\eta}] \text{ s.t. } E'\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}\}\}.$$

Notice that the inner problem depends on $\bar{\boldsymbol{\mu}}$, bu not on $\bar{\boldsymbol{u}}$. In other words, for any given $\boldsymbol{\eta}$, every $\bar{\boldsymbol{\mu}} \in \bar{M}$ is can be associated with a unique $\boldsymbol{\mu} \in M$. Hence, intuitively, $\bar{\boldsymbol{\mu}}$ conveys the same information about $\bar{\boldsymbol{u}}$ as the full state-action frequency. This also implies that the dual relationship between $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{u}}$ is preserved. To see this, let $\boldsymbol{\mu}(\bar{\boldsymbol{\mu}})$ denote the solution of the inner problem, and $\bar{\boldsymbol{w}}^*(\bar{\boldsymbol{\mu}})$ denote its value. It follows from Theorem 2 that $\bar{\boldsymbol{u}}$ rationalizes $\bar{\boldsymbol{\mu}}$ if and only if $\bar{\boldsymbol{\mu}}$ solves

$$\max_{\bar{\boldsymbol{\mu}} \in \bar{M}} [\bar{\boldsymbol{\mu}} \cdot \bar{\boldsymbol{u}} - \bar{\boldsymbol{w}}^*(\bar{\boldsymbol{\mu}})] \tag{25}$$

and that this is the case if and only if $\bar{\boldsymbol{u}}$ solves

$$\max_{\bar{\boldsymbol{u}} \in \mathbb{R}^Z} [\bar{\boldsymbol{\mu}} \cdot \bar{\boldsymbol{u}} - \bar{\boldsymbol{w}}(\bar{\boldsymbol{u}})]$$

where $\bar{\boldsymbol{w}}(\bar{\boldsymbol{u}})$ denotes the value of Problem 25. In turn, duality implies that all results on identification and inversion generalize to this case, as stated in remaining of this Section.

6.2.1 Identification

In terms of identification, one has that $\bar{\boldsymbol{\mu}}$ identifies the average payoffs associated with every feedback, up to a constant.

Proposition 15. *For every $\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}} \in \mathbb{R}^Z$, we have $\bar{\boldsymbol{\mu}}(\bar{\boldsymbol{u}}) = \bar{\boldsymbol{\mu}}(\bar{\boldsymbol{v}})$ if and only if there exists some constant $k \in \mathbb{R}$ such that*

$$\bar{\boldsymbol{\nu}} \cdot \bar{\boldsymbol{u}} = \bar{\boldsymbol{\nu}} \cdot \bar{\boldsymbol{v}} + k \text{ for all } \bar{\boldsymbol{\nu}} \in \bar{M} \ . \tag{26}$$

Letting

$$\bar{M}_0 = \{E' \boldsymbol{\mu} : \ \boldsymbol{\mu} \in M_0\}$$

be the set of the average payoffs of all pure feedbacks (that is, of all feedbacks associated with deterministic policies) it can be shown that $\bar{M}_0$ corresponds to the set of extreme points of $\bar{M}$, hence Condition 26 is equivalent to

$$\bar{M}_0' \bar{\boldsymbol{u}} = \bar{M}_0' \bar{\boldsymbol{v}} + k.$$

As a consequence of Proposition 15, similarly to Section 4, the degree of under-identification of $\bar{\boldsymbol{u}}$ is a function of $\dim \bar{M}$. In particular, $\bar{\boldsymbol{\mu}}$ identifies an affine set of average payoffs of dimension $|Z| - \dim \bar{M}$, and in order to identify $\bar{\boldsymbol{u}}$ the analyst must impose at least $|Z| - \dim \bar{M} + 1$ additional linear restrictions. Another consequence is that the results on the identifying power of linear restrictions extend directly as well. To state them formally, similarly to Section 4, if C is a matrix with columns in $\mathbb{R}^{|Z|}$ say that $C' \bar{\boldsymbol{u}} = 0$ identifies $\bar{\boldsymbol{u}}$ if, for all $\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}} \in \mathbb{R}^{|Z|}$, we have $\bar{\boldsymbol{u}} = \bar{\boldsymbol{v}}$ whenever $C' \bar{\boldsymbol{u}} = C' \bar{\boldsymbol{v}}$ and $\bar{\boldsymbol{\mu}}(\bar{\boldsymbol{u}}) = \bar{\boldsymbol{\mu}}(\bar{\boldsymbol{v}})$, and that $\bar{\boldsymbol{u}} = C\theta$ identifies $\bar{\boldsymbol{u}}$ if, for all $\theta, \delta \in \mathbb{R}^{|C|}$, we have $\theta = \delta$ whenever $\bar{\boldsymbol{\mu}}(C\theta) = \bar{\boldsymbol{\mu}}(C\delta)$. Then, replacing $\boldsymbol{u}$ with $\bar{\boldsymbol{u}}$, M with $\bar{M}$ and M_0 with $\bar{M}_0$ in the statement of Proposition 7, the result still holds. Given this,

the under-identification result of Proposition 8 generalizes as well. Formally, if C is such that $C'\bar{\boldsymbol{u}} = 0$ (resp. $\bar{\boldsymbol{u}} = C\theta$) identifies $\bar{\boldsymbol{u}}$ and $|C| = |Z| - \dim\bar{M} + 1$ (resp. $\dim\bar{M} - 1$) then, for every $\bar{\boldsymbol{\mu}} \in \bar{M}$ such that $\bar{\boldsymbol{\mu}}(z) > 0$ for all z, there exists $\bar{\boldsymbol{u}}$ such that $C'\bar{\boldsymbol{u}} = 0$ (resp. $\bar{\boldsymbol{u}} = C'\theta$ for some $\theta \in \mathbb{R}^{|C|}$) and $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}(\bar{\boldsymbol{u}})$.

Example 1 (continued) $\bar{\boldsymbol{u}}$ is identified up to a constant whenever there exist $|X|$ linearly independent stationary distributions over states associated with some policy.

Example 2 (continued) For every $1 \leq n \leq |X|$, let $\bar{\boldsymbol{\mu}}^n$ be the feedback associated with the deterministic policy that prescribes to replace the engine at x_m if and only if $n \leq m$, and let $\bar{\boldsymbol{\mu}}^{|X|+1}$ be the feedback associated with the deterministic policy that prescribes to never replace the engine. Letting $B = \{\bar{\boldsymbol{\mu}}^n : \ n = 1, ..., |X| + 1\}$, it is easy to see that B is linearly independent. Hence $\dim\bar{M} = |X| + 1$, implying that RC and $c = (c(x))_{x \in X}$ are identified up to a constant from $\bar{\boldsymbol{\mu}}$.

6.2.2 Inversion

Another consequence of duality is that the inversion procedure described in Section 5.1.2 generalizes directly as well, and progress bounds can be established similarly to Proposition 12. Formally, if C is such that $\bar{\boldsymbol{u}} = C\theta$ identifies $\bar{\boldsymbol{u}}$ then, for any given $\bar{\boldsymbol{\mu}} \in \bar{M}$ such that $\bar{\boldsymbol{\mu}}(z) > 0$ for all z, the problem

$$\max_{\theta \in \mathbb{R}^{|C|}} [\bar{\boldsymbol{\mu}} \cdot C\theta - \bar{\boldsymbol{w}}(C\theta)]$$

admits a unique solution θ^*. Again, this can computed for instance through the gradient method, which, for a given constant $\gamma > 0$ and an arbitrary starting point $\theta^0 \in \mathbb{R}^{|C|}$, generates the iterates

$$\theta^{n+1} = \theta^n + \gamma C'[\bar{\boldsymbol{\mu}} - \bar{\boldsymbol{\mu}}(C\theta^n)] \text{ for every } n \geq 0.$$

It can be shown that, if E is regular, for γ small enough, θ^n converges monotonically to θ^*, and progress bounds can be established similarly to Proposition 12. If $\dim\bar{M} - 1$ then θ^* is (the unique vector of parameters) such that $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}(C\theta^*)$. Otherwise, θ^* minimizes $D_{\bar{\boldsymbol{w}}^*}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}(C\theta))$ with respect to θ. Hence this procedure can be used to compute a vector $\bar{\boldsymbol{u}}^*$ such that $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}(\bar{\boldsymbol{u}}^*)$, so that θ can be estimated via constrained least squares, replacing $\nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ with $\bar{\boldsymbol{u}}^*$, $\mathbb{R}^{|A||X|}$ with $\mathbb{R}^{|Z|}$, and B with a linear basis of $\bar{M}$

in the treatment of Section 5.2.1, or to compute a Bregman projection estimator as in Section 5.2.2.

References

Eitan Altman and Adam Shwartz. *Non-stationary policies for controlled Markov Chains*. Technion-IIT, Department of Electrical Engineering, 1987.

Takeshi Amemiya. *Advanced econometrics*. Harvard university press, 1985.

Aristotle Arapostathis, Vivek S Borkar, Emmanuel Fernández-Gaucherand, Mrinal K Ghosh, and Steven I Marcus. Discrete-time controlled markov processes with average cost criterion: a survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.

Arindam Banerjee, Xin Guo, and Hui Wang. On the optimality of conditional expectation as a bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, 2005a.

Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, Joydeep Ghosh, and John Lafferty. Clustering with bregman divergences. *Journal of machine learning research*, 6(10), 2005b.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, pages 841–890, 1995.

Steven Berry, Amit Gandhi, and Philip Haile. Connected substitutes and invertibility of demand. *Econometrica*, 81(5):2087–2111, 2013.

Dimitri P Bertsekas. *Dynamic programming and optimal control*, volume 2. Athena scientific Belmont, MA, 1995.

David Blackwell. Discrete dynamic programming. *The Annals of Mathematical Statistics*, pages 719–726, 1962.

Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3): 200–217, 1967.

Stephen V Cameron and James J Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political economy*, 106(2):262–333, 1998.

Khai Xiang Chiong, Alfred Galichon, and Matt Shum. Duality in dynamic discrete-choice models. *Quantitative Economics*, 7(1):83–115, 2016.

Grace E Cho and Carl D Meyer. Comparison of perturbation bounds for the stationary distribution of a markov chain. *Linear Algebra and its Applications*, 335(1-3):137–150, 2001.

Cyrus Derman. Finite state markovian decision processes. 1970.

Awi Federgruen, PJ Schweitzer, and Hendrik Cornelis Tijms. Contraction mappings underlying undiscounted markov decision problems. *Journal of Mathematical Analysis and Applications*, 65(3):711–730, 1978.

Alfred Galichon and Bernard Salanié. Cupid's invisible hand: Social surplus and identification in matching models. *Working paper*, 2020.

Arie Hordijk and Lodewijk CM Kallenberg. Constrained undiscounted stochastic dynamic programming. *Mathematics of Operations Research*, 9(2):276–289, 1984.

V. Joseph Hotz and Robert A. Miller. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60(3):497–529, 1993.

Shiro Ishikawa. Fixed points and iteration of a nonexpansive mapping in a banach space. *Proceedings of the American Mathematical Society*, 59:65–71, 1976.

Sham Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. *Unpublished Manuscript*, 2009.

Lodewijk CM Kallenberg. Survey of linear programming for standard and nonstandard markovian control problems. part ii: Applications. *Zeitschrift für Operations Research*, 40(2):127–143, 1994a.

Lodewijk CM Kallenberg. Survey of linear programming for standard and nonstandard markovian control problems. part i: Theory. *Zeitschrift für Operations Research*, 40(1):1–42, 1994b.

Harold W Kuhn. Extensive games and the problem of information. *Annals of Mathematics*, pages 193–216, 1953.

R Duncan Luce. *Individual choice behavior: A theoretical analysis*. John Wiley and sons, 1959.

Thierry Magnac and David Thesmar. Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816, 2002.

Charles F Manski. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542, 1993.

Daniel McFadden. Modeling the choice of residential location. *Transportation Research Record*, (673), 1978.

Loren Platzman. Improved conditions for convergence in undiscounted markov renewal programming. *Operations research*, 25(3):529–533, 1977.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 1970.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5): 999–1033, 1987.

John Rust. Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143, 1994.

Xiaoxia Shi, Matthew Shum, and Wei Song. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Econometrica*, 86(2):737–761, 2018.

Christopher R Taber. Semiparametric identification and heterogeneity in discrete choice dynamic programming models. *Journal of econometrics*, 96(2):201–229, 2000.

Douglas J White. Dynamic programming, markov chains, and the method of successive approximations. *Journal of Mathematical Analysis and Applications*, 6(3):373–376, 1963.

A Preliminaries

This Section provides a brief review of some notions and results used in the paper. I keep an informal style, but I refer to Rockafellar [1970] for an extensive treatment of the subject. Let $g:\ \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a continuous and convex function not identically equal to $+\infty$. The domain of g is defined as the set of all $z \in \mathbb{R}^N$ such that $g(z) \neq +\infty$. Its sub-gradient at $z \in \mathbb{R}^N$, denoted by $\nabla g(z) \subseteq \mathbb{R}^N$, is the set of all vectors $y \in \mathbb{R}^N$ such that

$$g(z') - g(z) \geq y \cdot [z' - z] \text{ for all } z' \in \mathbb{R}^N \ .$$

$\nabla g(z)$ is never empty in the relative interior of the domain of g. Moreover, g is differentiable at z if and only if $\nabla g(z)$ is a singleton, and its unique element is the gradient of g at z. If this is the case, abusing notation, I will also denote the gradient of g at z by $\nabla g(z)$.

The convex conjugate of g is the function g^* defined by

$$g^*(z) = \sup_{y \in \mathbb{R}^N} [z \cdot y - g(y)] \text{ for all } z \in \mathbb{R}^N \ .$$

A fundamental result is that strong duality holds, that is, we have $g^{**} = g$, or

$$g(z) = \sup_{y \in \mathbb{R}^N} [z \cdot y - g^*(y)] \text{ for all } z \in \mathbb{R}^N \ .$$

g^* is always convex and such that

$$g(y) + g^*(z) \geq z \cdot y \text{ for all } y, z \in \mathbb{R}^N$$

and the above inequality holds as an equality if and only if $z \in \nabla g(y)$, or, equivalently, $y \in \nabla g^*(z)$. Moreover, the following result holds, which I state formally for future reference.

Lemma 3. *For a convex function g the following are equivalent:*

i) g is differentiable

ii) g is continuously differentiable

iii) g^* is strictly convex

Another important result is the conjugacy between smoothness and strong convexity. Formally, if $\|\cdot\|$ is a norm on $\mathbb{R}^N$, say that

1. g is *strongly convex* with respect to $\|\cdot\|$ if there exists a scalar $K > 0$ such that, for every y, z in the relative interior of the domain of g and every $\lambda \in (0, 1)$, we have

$$g(\lambda y + (1-\lambda)z) \leq \lambda f(y) + (1-\lambda) f(z) - K\lambda(1-\lambda) \|z - y\|^2 \ .$$

2. g is *smooth* with respect to $\|\cdot\|$ if it is differentiable in the relative interior of its domain, and there exists a scalar $L > 0$ such that, for every y, z in the relative interior of the domain of g, we have

$$g(z) \leq g(y) + \nabla g(y) \cdot (z - y) + L \|z - y\|^2 \ .$$

Notice that, since all norms on $\mathbb{R}^N$ are equivalent, g is strongly convex with respect to some norm if and only if it is strongly convex with respect to all norms. And similarly, g is smooth with respect to some

norm if and only if it is smooth with respect to all norms. For this reason, in what follows I will omit the reference to a particular norm, and simply say that f is strongly convex, or smooth.

Lemma 4. *g is strongly convex if and only if g^* is smooth.*

In words, strong convexity and smoothness are conjugate properties.. A proof of this result is provided for instance by Kakade et al. [2009, Theorem 6].

B Proofs

B.1 Proof of Theorem 2

B.1.1 Proof that w and w^* are strictly convex and continuously differentiable

Under Assumption 2, w is differentiable (see Lemma 1 in Shi et al. 2018). Hence Lemma 3 implies that w^* is strictly convex. To see that $\boldsymbol{w}^*$ is strictly convex, take $\boldsymbol{\mu}, \boldsymbol{\nu} \in M$ and $\alpha \in (0,1)$. For every state x, let $\lambda(x) \equiv \alpha \boldsymbol{\mu}_X(x)/[\alpha \boldsymbol{\mu}_X(x) + (1-\alpha)\boldsymbol{\nu}_X(x)]$, and notice that we have[23]

$$\boldsymbol{\sigma}^{\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}}(x) = \lambda(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) + [1-\lambda(x)]\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x).$$

[23] That is, we have

$$\begin{aligned}\boldsymbol{\sigma}^{\alpha\mu+(1-\alpha)\nu}(a,x) \equiv & \frac{\alpha\boldsymbol{\mu}(a,x) + (1-\alpha)\boldsymbol{\nu}(a,x)}{\alpha\boldsymbol{\mu}_X(x) + (1-\alpha)\boldsymbol{\nu}_X(x)} = \frac{\boldsymbol{\alpha}\mu(x)}{\alpha\boldsymbol{\mu}_X(x) + (1-\alpha)\boldsymbol{\nu}_X(x)}\frac{\boldsymbol{\mu}(a,x)}{\boldsymbol{\mu}_X(x)} + \frac{(1-\alpha)\boldsymbol{\nu}_X(x)}{\alpha\boldsymbol{\mu}_X(x) + (1-\alpha)\boldsymbol{\nu}_X(x)}\frac{\boldsymbol{\nu}(a,x)}{\boldsymbol{\nu}_X(x)} \\ = & \lambda(x)\frac{\boldsymbol{\mu}(a,x)}{\boldsymbol{\mu}_X(x)} + [1-\lambda(x)]\frac{\boldsymbol{\nu}(a,x)}{\boldsymbol{\nu}_X(x)} = \lambda(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) + [1-\lambda(x)]\boldsymbol{\sigma}^{\boldsymbol{\nu}}(a,x)\end{aligned}$$

for every state x and action a.

By convexity of w^*, this implies

$$\begin{aligned}
\boldsymbol{w}^*(\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}) =& \sum_x [\alpha\boldsymbol{\mu}_X(x)+(1-\alpha)\boldsymbol{\nu}_X(x)]w^*(\boldsymbol{\sigma}^{\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}}(x)) \\
=& \sum_x [\alpha\boldsymbol{\mu}_X(x)+(1-\alpha)\boldsymbol{\nu}_X(x)]w^*(\lambda(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)+[1-\lambda(x)]\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)) \\
\leq& \sum_x [\alpha\boldsymbol{\mu}_X(x)+(1-\alpha)\boldsymbol{\nu}_X(x)]\{\lambda(x)w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x))+[1-\lambda(x)]w^*(\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x))\} \\
=& \sum_x [\alpha\boldsymbol{\mu}_X(x)w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x))+(1-\alpha)\boldsymbol{\nu}_X(x)]w^*(\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x))] \\
=& \alpha\boldsymbol{w}^*(\boldsymbol{\mu})+(1-\alpha)\boldsymbol{w}^*(\boldsymbol{\nu})
\end{aligned}$$

Moreover, strict convexity of w^* implies that the inequality is strict whenever $\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) \neq \boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)$ for some x such that $\boldsymbol{\mu}_X(x)+\boldsymbol{\nu}_X(x) > 0$, which is the case whenever $\boldsymbol{\mu} \neq \boldsymbol{\nu}$. This shows that $\boldsymbol{w}^*$ is strictly convex. This implies that the function

$$\tilde{\boldsymbol{w}}^*(\boldsymbol{\mu}) = \begin{cases} \boldsymbol{w}^*(\boldsymbol{\mu}) & \text{if } \boldsymbol{\mu} \in M \\ +\infty & \text{otherwise} \end{cases}$$

is also strictly convex. Since $\boldsymbol{w} = (\tilde{\boldsymbol{w}}^*)^*$, this implies that $\boldsymbol{w}$ is convex and, by Lemma 3, continuously differentiable. It remains to show that $\boldsymbol{w}^*$ is continuously differentiable. By Lemma 3, it suffices to show that it is differentiable. I prove this in the following Lemma, which I state separately for future reference.

Lemma 5. *$\boldsymbol{w}^*$ is differentiable and such that*

$$E_F \max_a [\frac{d\boldsymbol{w}^*(\mu)}{d\boldsymbol{\mu}(a,x)} + \epsilon(a)] = 0$$

and

$$\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) = Pr_F[\frac{d\boldsymbol{w}^*(\mu)}{d\boldsymbol{\mu}(a,x)} + \epsilon(a) = \max_{a' \in A} \frac{d\boldsymbol{w}^*(\mu)}{d\boldsymbol{\mu}(a,x)} + \epsilon(a')]$$

for every a, x.

Proof. It is well-known that w^* is such that, for every $\sigma \in \Delta A$ and every $u, v \in \nabla w^*(\sigma)$, there exists

some scalar $k \in \mathbb{R}$ such that $v = u + k$. This implies that $u - \sigma \cdot u = v - \sigma \cdot v$, so that one can define

$$\nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma) \equiv u(a) - \sigma \cdot u$$

for an arbitrary $u \in \nabla w^*(\sigma)$. I now show that the following equivalence holds for any $u \in \mathbb{R}^A$ and $\sigma \in \Delta A$:

$$u = w^*(\sigma) + \nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma) \Leftrightarrow u \in \nabla w^*(\sigma) \text{ and } w(u) = 0 \ .$$

To see this note that, if $u = w^*(\sigma) + \nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma)$, then $u \in \nabla w^*(\sigma)$. By Theorem 1, this implies $w(u) = \sigma \cdot u - w^*(\sigma)$, so that

$$w(u) = \sigma \cdot u - u + \nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma) = 0 \ .$$

Conversely, if $u \in \nabla w^*(\sigma)$ and $w(u) = 0$ then, by Theorem 1, $w^*(\sigma) = \sigma \cdot u - w(u) = \sigma \cdot u$. This implies

$$w^*(\sigma) + \nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma) = \sigma \cdot u + \nabla w^*(\sigma) - \sigma \cdot \nabla w^*(\sigma) = u \ .$$

Given this notice that, if $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$, then $\boldsymbol{u} \in \nabla \boldsymbol{w}^*(\boldsymbol{\mu})$ if and only if

$$\boldsymbol{u}(a,x) = w^*(\boldsymbol{\sigma^\mu}(x)) + \frac{\mathrm{d}w^*(\boldsymbol{\sigma^\mu}(x))}{\mathrm{d}\sigma(a)} - \boldsymbol{\sigma^\mu}(x) \cdot \nabla w^*(\boldsymbol{\sigma^\mu}(x)) \text{ for all } a, x \ .$$

This shows that $\boldsymbol{w}^*$ is differentiable, and that

$$\boldsymbol{u}(x) \in \nabla w^*(\boldsymbol{\sigma^\mu}(x)) \text{ and } w(\boldsymbol{u}(x)) = 0 \text{ for all } x \ .$$

The statement then follows from Theorem 1. □

B.1.2 Proof the main part of the statement

The idea behind the result is simple: intuitively, in analogy with the static case, the quantity $-\boldsymbol{w}^*(\boldsymbol{\mu})$ can be thought as the maximum expected average utility from shocks that is achievable by matching shocks

to actions in each state in a way that is consistent with $\boldsymbol{\mu}$. Formally, if $\boldsymbol{\mu} \in M$ we have

$$-\boldsymbol{w}^*(\boldsymbol{\mu}) = \sup_{\boldsymbol{\pi}} \sum_x \boldsymbol{\mu}_X(x) \mathrm{E}_F[\epsilon(\boldsymbol{\pi}(x,\epsilon))] \text{ s.t. } \Pr_F[\boldsymbol{\pi}(x,\epsilon) = a] = \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) \text{ for all } a, x$$

where the supremum is taken with respect to all stationary policies $\boldsymbol{\pi}$. This follows immediately from Proposition 4, since the above problem is equivalent to

$$\sum_x \boldsymbol{\mu}_X(x) \{ \sup_{\pi \colon \mathbb{R}^A \to A} \mathrm{E}_F[\epsilon(\pi(\epsilon))] \text{ s.t. } \Pr_F[\pi(\epsilon) = a] = \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) \text{ for all } a \} \ .$$

Hence the problem of choosing an optimal policy can be split into two nested problems: an outer problem in which an optimal stationary measure is chosen, and an inner problem in which an optimal policy $\boldsymbol{\pi}$ is chosen in order to maximize the expected utility arising from shocks, subject to it being consistent with $\boldsymbol{\mu}$. Formally we have

$$\begin{aligned}
\max_{\boldsymbol{\pi}} \boldsymbol{w}_{\boldsymbol{\pi}}(\boldsymbol{u}|x_0) &= \max_{\boldsymbol{\pi}} \boldsymbol{\mu}(\boldsymbol{\pi}, x_0) \cdot \boldsymbol{u} + \sum_x \boldsymbol{\mu}_X(x|\boldsymbol{\pi}, x_0) \mathrm{E}_F[\epsilon(\boldsymbol{\pi}(x,\epsilon))] \\
&= \max_{\boldsymbol{\mu} \in M_+} \{ \max_{\boldsymbol{\pi}} \boldsymbol{\mu} \cdot \boldsymbol{u} + \sum_x \boldsymbol{\mu}_X(x) \mathrm{E}_F[\epsilon(\boldsymbol{\pi}(x,\epsilon))] \} \text{ s.t. } \boldsymbol{\mu}(\boldsymbol{\pi}, x_0) = \boldsymbol{\mu} \} \\
&= \max_{\boldsymbol{\mu} \in M_+} \boldsymbol{\mu} \cdot \boldsymbol{u} + \{ \max_{\boldsymbol{\pi}} \sum_x \boldsymbol{\mu}_X(x) \mathrm{E}_F[\epsilon(\boldsymbol{\pi}(x,\epsilon))] \} \text{ s.t. } \Pr_F[\boldsymbol{\pi}(x,\epsilon) = a] = \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) \text{ for all } a, x \} \\
&= \max_{\boldsymbol{\mu} \in M_+} \boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}) = \max_{\boldsymbol{\mu} \in M} \boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu})
\end{aligned}$$

Hence, the second equality follows from the fact that every optimal policy must induce full-support choice probabilities. The third equality follows from the fact that, if $\boldsymbol{\mu} \in M_+$, then $\boldsymbol{\mu}(\boldsymbol{\pi}, x_0)$ does not depend on the initial state, while The fourth equality follows from previous result. This shows that

$$\boldsymbol{w}(\boldsymbol{u}) = \max_{\boldsymbol{\mu} \in M} \boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}) \ .$$

Hence it shows that $\boldsymbol{w}$ and the function

$$\tilde{\boldsymbol{w}}^*(\boldsymbol{\mu}) = \begin{cases} \boldsymbol{w}^*(\boldsymbol{\mu}) & \text{if } \boldsymbol{\mu} \in M \\ +\infty & \text{otherwise} \end{cases}$$

are convex conjugates. Hence the statement follows from Fenchel's duality Theorem.

B.2 Proof of Proposition 1

By making the constraints defining M explicit, Problem 10 writes as

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^{|A||X|}} \; & \boldsymbol{\mu}\cdot\boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}) \\ \text{s.t.} \quad & \textstyle\sum_a \boldsymbol{\mu}(a,x) = \sum_{a',x'} \boldsymbol{\mu}(a',x')T(x|a',x') \qquad \forall\, x \in X \\ & \textstyle\sum_a \boldsymbol{\mu}(a,x) = 1 \end{aligned}$$

By Theorem 2, this problem is strictly convex and differentiable. Let $V \in \mathbb{R}^X$ and $\boldsymbol{w} \in \mathbb{R}$ denote the multipliers of the first set of constraints and of the last constraint, respectively. By taking the first order conditions, it is easy to see that $\boldsymbol{\mu}, V, \boldsymbol{w}$ is an optimal dual pair if and only if it satisfies

$$\boldsymbol{u}(a,x) + T(a,x)\cdot V - V(x) - \boldsymbol{w} = \frac{\mathrm{d}\boldsymbol{w}^*(\mu)}{\mathrm{d}\boldsymbol{\mu}(a,x)} \text{ for every } a, x \ .$$

The result then follows from Lemma 5.

B.3 Proof of Proposition 2

Proof. The quantity $(1-\beta)V^\beta(x)$ is a weighted average of expected future payoffs. The expected payoff in each period must be greater than

$$\underline{u} \equiv \min_{a,x} \boldsymbol{u}(a,x) + \mathrm{E}_F[\min_a \epsilon(a)]$$

and smaller than

$$\bar{u} \equiv \max_{a,x} \boldsymbol{u}(a,x) + \mathrm{E}_F[\max_a \epsilon(a)] \ .$$

Hence $\underline{u} \leq (1-\beta)V^\beta(x) \leq \bar{u}$ for every every state x. Similarly, one can find bounds on the differences $V^\beta(x) - V^\beta(x')$ that are independent of β. To see this, notice that $(1-\beta^t)V^\beta(x)$ can be bounded for

every integer t. Indeed, we have

$$(1-\beta^t)V^\beta(x) = \beta(1-\beta^{t-1})V^\beta(x) + (1-\beta)V^\beta(x)$$

Therefore, if $(1-\beta^{t-1})V^\beta(x)$ can be bounded, so can $(1-\beta^t)V^\beta(x)$, and the result follows by induction. Now let $\boldsymbol{\sigma}$ be any full-support system of conditional choice probabilities, and consider the stochastic process under $\boldsymbol{\sigma}$ starting in state x. Let $\tilde{t} \geq 1$ be the first period in which the process visits state x' (a random variable). By Accessibility it follows that $\sum_{t=1}^{\infty} \Pr[\tilde{t}=t] = 1$ and that $\mathrm{E}[\tilde{t}] < \infty$. For each period t, letting u_t be the per-period payoff at t (a random variable), we have that $\beta^\tau \mathrm{E}[u_\tau | \tilde{t}=t] \geq \min(\underline{u}, 0)$ for every $0 \leq \tau \leq t$. Hence

$$V^\beta(x) \geq \sum_{t=1}^{\infty} \Pr[\tilde{t}=t][t\min(\underline{u},0) + \beta^t V^\beta(x')] = \min(\underline{u},0)\mathrm{E}[\tilde{t}] + \sum_{t=1}^{\infty} \Pr[\tilde{t}=t]\beta^t V^\beta(x') \ .$$

Subtracting $V^\beta(x')$ from both sides yields

$$V^\beta(x) - V^\beta(x') \geq \min(\underline{u},0)\mathrm{E}[\tilde{t}] - \sum_{t=1}^{\infty} \Pr[\tilde{t}=t](1-\beta^t)V^\beta(x') \ .$$

Hence $V^\beta(x) - V^\beta(x')$ is bounded below. Reversing the roles of x and x' shows that $V^\beta(x) - V^\beta(x')$ is also bounded above.

From boundedness it follows that, for any sequence $(\beta_n)_{n\geq 0} \subseteq (0,1)$ of discount factors such that $\lim_{n\to\infty} \beta_n = 1$, there is a sub-sequence $(\beta_m)_{m\geq 0} \subseteq (0,1)$ such that $((1-\beta_m)V^{\beta_m}(x))_{m\geq 0}$ and $(V^{\beta_m}(x) - V^{\beta_m}(x'))_{m\geq 0}$ converge for every x, x'. Equation 6 then implies that, for every x, x',

$$\lim_{m\to\infty}(1-\beta_m)V^{\beta_m}(x) = \boldsymbol{w} \text{ and } \lim_{m\to\infty} V^{\beta_m}(x) - V^{\beta_m}(x') = V(x) - V(x')$$

for some $V, \boldsymbol{w}$ satisfying the recursive Equations 3. The result then follows from Corollary 2. □

B.4 Proof of Proposition 3

B.4.1 A norm on $\mathbb{R}^X$

In what follows I am going to work with the norm $\|\cdot\|$ defined on $\mathbb{R}^X$ by

$$\|V\| = \max(\max_{x\in X} V(x), 0) - \min(\min_{x\in X} V(x), 0)$$

Below I show that $\|\cdot\|$ satisfies all properties defining norms.

Proposition 16. *For every $V, W \in \mathbb{R}^X$ and $k \in \mathbb{R}$ we have:*

1. $\|V\| \geq 0$
2. $\|V\| = 0$ if and only if $V = 0_{|X|}$
3. $\|kV\| = |k| \, \|V\|$
4. $\|V + W\| \leq \|V\| + \|W\|$.

Proof. 1 follows from $\max(\max_{x\in X} V(x), 0) \geq 0$ and $\min(\min_{x\in X} V(x), 0) \leq 0$. 2 follows from the fact that a is the null vector if and only if $\max_{x\in X} V(x) = \min_{x\in X} V(x) = 0$, and this is the case if and only if $\|V\| = 0$. To see that 3 holds note that if k is positive then

$$\max(\max_{x\in X} kV(x), 0) = |k| \max(\max_{x\in X} V(x), 0) \text{ and } \min(\min_{x\in X} kV(x), 0) = |k| \min(\min_{x\in X} a(x), 0)$$

while if k is negative then

$$\max(\max_{x\in X} kV(x), 0) = -|k| \min(\min_{x\in X} V(x), 0) \text{ and } \min(\min_{x\in X} kV(x), 0) = -|k| \max(\max_{x\in X} V(x), 0) \ .$$

Finally, 4 follows from

$$\max(\max_{x\in X}(V(x) + W(x)), 0) \leq \max(\max_{x\in X} V(x) + \max_{x\in X} W(x), 0) \leq \max(\max_{x\in X} V(x), 0) + \max(\max_{x\in X} W(x), 0)$$

and

$$\min(\min_{x\in X}(V(x)+W(x)),0)\geq \min(\min_{x\in X}V(x)+\min_{x\in X}W(x),0)\geq \min(\min_{x\in X}V(x),0)+\min(\min_{x\in X}W(x),0)\ .$$

□

B.4.2 Fixed points of Γ

Lemma 6. *Fix an arbitrary state x_0 and define the relative value operator Γ as in Section 1.3. Γ admits a unique fixed point.*

Proof. Take $V, \boldsymbol{w}$ satisfying the recursive Equations in 3 - notice that, by Proposition 1, such $V, \boldsymbol{w}$ always exist, otherwise Problem 10 would have no solution. To see that Γ has a fixed point V^*, let $V^*(x) = V(x) - V(x_0)$ for every x. Then it is easy to see that $\Gamma V^* = V^*$. I show that such fixed point is unique. For every $V \in \mathbb{R}^X$, define the system of conditional choice probabilities $\boldsymbol{\sigma}^V$ and the transition matrix T^V of associated with V by

$$\boldsymbol{\sigma}^V(a,x) := \Pr_F[a = \arg\max_{a'\in A}[\boldsymbol{u}(a',x)+T(a',x)\cdot V+\epsilon(a')]] \text{ and } T^V(x'|x) := \sum_a \boldsymbol{\sigma}^V(a,x)T(x'|a,x)$$

and the constant w^V by

$$w^V := \mathrm{E}_F[\max_{a\in A}[\boldsymbol{u}(a,x_0)+T(a,x_0)\cdot V+\epsilon(a)]]\ .$$

In what follows, I will treat $\boldsymbol{\sigma}^V$ and $\boldsymbol{u}$ as the $|X|\times|A|$-dimensional matrices whose generic x,a-th coordinates are given by $\boldsymbol{\sigma}^V(a,x)$ and $\boldsymbol{u}(a,x)$, respectively. I will also treat T as a $|X|\times|X|$-dimensional matrix whose generic x,x'-th coordinate is given by $T(x'|x)$. Finally, I let $w^*(\boldsymbol{\sigma}^V)$ denote the $|X|$-dimensional vector with generic x-th coordinate given by $w^*(\boldsymbol{\sigma}^V(x))$. Suppose that U and V are both fixed points of Γ. Then we must have $U(x_0) = V(x_0) = 0$ and

$$U = \boldsymbol{\sigma}^U\boldsymbol{u}' + T^U U - w^*(\boldsymbol{\sigma}^U) - w^U \geq \boldsymbol{\sigma}^V\boldsymbol{u}' + T^V U - w^*(\boldsymbol{\sigma}^V) - w^U$$
$$V = \boldsymbol{\sigma}^V\boldsymbol{u}' + T^V V - w^*(\boldsymbol{\sigma}^V) - w^V \geq \boldsymbol{\sigma}^U\boldsymbol{u}' + T^U V - w^*(\boldsymbol{\sigma}^U) - w^V$$

and therefore

$$T^U(V-U)+w^U-w^V \leq V-U \leq T^V(V-U)+w^U-w^V \ . \tag{27}$$

Since $\boldsymbol{\sigma}^V$ and $\boldsymbol{\sigma}^U$ have full support then, by Accessibility, there exists an integer K and a scalar $\epsilon \in (0,1)$ such that the (x,x_0)-th elements of $T^{U,K} \equiv \Pi_{k=1}^K T^U$ and $T^{V,K} \equiv \Pi_{k=1}^K T^V$ are greater than ϵ. Iteration of the above inequalities gives

$$T^{U,K}(V-U)+K(w^U-w^V) \leq V-U \leq T^{V,K}(V-U)+K(w^U-w^V)$$

Hence, using the fact that $U(x_0)=V(x_0)=0$, we get

$$\begin{aligned}
\max[\max_x[V(x)-U(x),0]] &\leq (1-\epsilon)\max[\max_x[V(x)-U(x),0]]+K(w^U-w^V) \\
\min[\min_x[V(x)-U(x)],0] &\geq (1-\epsilon)\min[\min_x[V(x)-U(x)],0]+K(w^U-w^V).
\end{aligned}$$

Subtracting these two inequalities gives

$$\|V(x)-U(x)\| \leq (1-\epsilon)\,\|V(x)-U(x)\|$$

which implies that $\|V(x)-U(x)\|=0$, completing the proof. □

Notice that this implies that the vector of relative valuations associated with the undiscounted problem is unique up to a constant. I state this below for future reference.

Corollary 2. *Suppose that $\boldsymbol{w}, V$ and $\boldsymbol{w}', V'$ both satisfy the recursive Equations 3. Then $\boldsymbol{w}'=\boldsymbol{w}$ and there exists a scalar $k \in \mathbb{R}$ such that $V'=V+k$.*

B.4.3 Proof of Proposition 3

Fix a state x_0 and a scalar $\alpha \in (0,1)$, and define Γ^α as in Section 1.3. The fact that Γ^α has a unique fixed point V^* follows from Lemma 6, since the sets of fixed points of Γ and Γ^α coincide. To show that $(V^k)_{k\geq 0}$ converges to V^* I exploit the following Lemma, which is a direct consequence of Ishikawa [1976, Theorem 1], a more general convergence result for non-expansive mappings.

Lemma 7. *Suppose that i)* Γ *is non-expansive according to* $\|\cdot\|$ *and ii) there exists a compact subset* D *of* $\mathbb{R}^X$ *such that* $V^0 \in D$ *and* Γ *maps* D *into* D*. Then* $(V^k)_{k\geq 0}$ *converges to* V^**.*

I start by showing that Γ is non-expansive according to $\|\cdot\|$. Following the same reasoning yielding to Equation 27, it is easy to see that for every $U, V \in \mathbb{R}^X$ we have

$$T^U(V-U) + w^U - w^V \leq \Gamma V - \Gamma U \leq T^V(V-U) + w^U - w^V \ .$$

From this we obtain

$$\min_{x\in X}[V(x)-U(x)] + w^U - w^V \leq \Gamma V - \Gamma U \leq \max_{x\in X}[V(x)-U(x)] + w^U - w^V$$

hence

$$\min_{x\in X}[\Gamma V(x) - \Gamma U(x)] \geq \min_{x\in X}[V(x)-U(x)] + w^U - w^V$$

$$\max_{x\in X}[\Gamma V(x) - \Gamma U(x)] \leq \max_{x\in X}[V(x)-U(x)] + w^U - w^V \ .$$

Subtracting these two equations yields

$$\begin{aligned}\max_{x\in X}[\Gamma V(x) - \Gamma U(x)] - \min_{x\in X}[\Gamma V(x) - \Gamma U(x)] &\leq \max_{x\in X}[V(x)-U(x)] - \min_{x\in X}[V(x)-U(x)] \\ &\leq \max\{\max_{x\in X}[V(x)-U(x)], 0\} - \min\{\min_{x\in X}[V(x)-U(x)], 0\}\end{aligned}$$

Since $\Gamma V(x_0) = \Gamma U(x_0)$ by construction, the left hand side equals $\|\Gamma V - \Gamma U\|$, while the right hand side defines $\|V - U\|$. Hence $\|\Gamma V - \Gamma U\| \leq \|V - U\|$ for every $U, V \in \mathbb{R}^X$.

To see that Condition ii) in previous Lemma is also satisfied, define $D \subseteq \mathbb{R}^X$ by

$$D = \{V : \|V - V^*\| \leq \|V_0 - V^*\|\} .$$

Clearly D is compact, and $V_0 \in D$ by construction. Moreover Γ maps D into D since it is non expansive with respect to $\|\cdot\|$.

Hence Conditions i) and ii) in previous Lemma are both satisfied, implying that and $V^k \to V^*$ as

$k \to \infty$.

B.5 Proof of Proposition 6

It remains to show the if direction of the statement. By Proposition 5, $(\boldsymbol{\mu}^{\mathbb{D}})_{\mathbb{D}\in\mathcal{D}}$ is rationalized by F if and only if there exists $\boldsymbol{u} \in \mathbb{R}^{|A||X|}$ such that, for every $\mathbb{D} \in \mathcal{D}$, there exist $k \in \mathbb{R}$ such that

$$\boldsymbol{\mu} \cdot \boldsymbol{u} = \boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}}) + k \ \forall \ \boldsymbol{\mu} \in M_0^{\mathbb{D}} \ . \tag{28}$$

Define $\boldsymbol{u}$ by $\boldsymbol{u}(a,x) = \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*})(a,x)$ for all a, x. Then Condition 28 holds trivially for $\mathbb{D} = \mathbb{D}^*$, with $k = 0$. For every $\mathbb{D} \neq \mathbb{D}^*$, fix an arbitrary $\boldsymbol{\nu} \in M_0^{\mathbb{D}}$, and let

$$k^{\mathbb{D}} \equiv \sum_{a,x} \boldsymbol{\nu}(a,x)[\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*})(a,x) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})(a,x)] \ .$$

Notice that Condition 16 is equivalent to

$$\forall \ \mathbb{D}, \mathbb{D}' \in \mathcal{D}, \ \forall \ \boldsymbol{\mu}, \boldsymbol{\nu} \in M_0^{\mathbb{D}} \cap M_0^{\mathbb{D}'} : \ (\boldsymbol{\mu} - \boldsymbol{\nu}) \cdot [\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}}) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}'})] = 0 \ .$$

Then for every $\boldsymbol{\mu} \in M_0^{\mathbb{D}}$ we have

$$\begin{aligned}
&\boldsymbol{\mu} \cdot [\boldsymbol{u} - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})] = \boldsymbol{\mu}[\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*}) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})] \\
&= (\boldsymbol{\mu} - \boldsymbol{\nu})[\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*}) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})] + \boldsymbol{\nu} \cdot [\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*}) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})] \\
&= \boldsymbol{\nu} \cdot [\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}^*}) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^{\mathbb{D}})] = k^{\mathbb{D}}.
\end{aligned}$$

That is, Condition 28 holds with $k = k^{\mathbb{D}}$. This completes the proof.

B.6 Proof of Lemma 1

Let D be the size of a maximal linearly independent subset of M, and B be the $|X| \times |A|\,|X|$-dimensional matrix whose coordinate in position $x, (a, x')$ is given by $T(x|a, x') - 1\{x = x'\}$. First, I show that $D = N \equiv |A|\,|X| - \text{rank}B$. To see this, notice that, for any $\boldsymbol{\mu} \in \mathbb{R}^{|A||X|}$, $\boldsymbol{\mu} \in M$ if and only if $B\boldsymbol{\mu} = 0$ and the coordinates of $\boldsymbol{\mu}$ are all positive and sum to one. Hence we must have $D \leq N$. To see that $D \geq N$, let

$\boldsymbol{\mu}^* \in D$ be an arbitrary stationary measure of full support. Since the solution space of the system $B\boldsymbol{\nu} = 0$ has dimension N, we can find $\boldsymbol{\nu}^1, ..., \boldsymbol{\nu}^{N-1} \in \mathbb{R}^{|A||X|}$ such that $B\boldsymbol{\nu}^n = 0$ for all n and $\{\boldsymbol{\nu}^1, ..., \boldsymbol{\nu}^{N-1}, \boldsymbol{\mu}^*\}$ is linearly independent. Take $\lambda \in \mathbb{R}$ large enough so that, for every n, $\boldsymbol{\mu}^n \equiv \boldsymbol{\nu}^n + \lambda\boldsymbol{\mu}^*$ has all coordinates strictly positive. $\{\boldsymbol{y}^1, ..., \boldsymbol{y}^{N-1}, \boldsymbol{\mu}^*\}$ is linearly independent since, for every $\alpha \in \mathbb{R}^N$, we have

$$\sum_{n=1}^{N-1} \alpha_n \boldsymbol{\mu}^n + \alpha_N \boldsymbol{\mu}^* = 0 \Leftrightarrow \sum_{n=1}^{N-1} \alpha_n \boldsymbol{\nu}^n + (\alpha_N + \lambda \sum_{n=1}^{N-1} \alpha_n)\boldsymbol{\mu}^* \Leftrightarrow \alpha = 0_N \ .$$

Defining $\tilde{\boldsymbol{\mu}}^n = \boldsymbol{y}^n / \sum_{a,x} \boldsymbol{\mu}^n(a,x)$ for every $n = 1, ..., N-1$ and $\tilde{\boldsymbol{\mu}}^N = \boldsymbol{\mu}^*$, we have that $\{\boldsymbol{\mu}^1, ..., \boldsymbol{\mu}^N\}$ is linearly independent. Hence $D = N$.

Given this, I now show that $\text{rank}B = |X| - 1$, which will then implies $D = |A|\,|X| - \text{rank}B = (|A| - 1)\,|X| + 1$. To see that $\text{rank}B \leq |X| - 1$, notice that the rows of B sum to the null vector, since

$$\sum_x [T(x|a,x') - 1\{x = x'\}] = \sum_x T(x|a,x') - 1 = 0 \text{ for every } a, x' \ .$$

To see that $\text{rank}B \geq |X| - 1$, suppose by way of contradiction that B has $|X| - 1$ linearly dependent rows. That is, there exist $|X| - 1$ distinct states $x_1, ..., x_{|X|-1}$ and a non-null vector $\alpha \in \mathbb{R}^{|X|-1}$ such that

$$\sum_{n=1}^{|X|-1} \alpha_n [T_a(x_n|x) - 1\{x = x_n\}] = 0 \text{ for every } a, x \ .$$

Take an arbitrary system $\boldsymbol{\sigma} \in (\Delta A)^X$ of conditional choice probabilities of full support, and let $T^{\boldsymbol{\sigma}} \in \mathbb{R}^{|X|\times|X|}$ be the associated matrix of state transition probabilities (i.e. $T^{\boldsymbol{\sigma}}(x'|x) = \sum_a \boldsymbol{\sigma}(a,x)T(x'|a,x)$ for every x, x'). We have

$$\sum_{n=1}^{|X|-1} \alpha_n [T^{\boldsymbol{\sigma}}(x_n|x) - 1\{x = x_n\}] = 0 \text{ for every } x \ .$$

Hence $T^{\boldsymbol{\sigma}}$ has $|X| - 1$ linearly dependent rows, that is, $\text{rank}T^{\boldsymbol{\sigma}} \leq |X| - 1$. This implies that $T^{\boldsymbol{\sigma}}$ admits at least two distinct stationary distributions (this can be seen through an argument similar to the one used in the proof that $D = |A|\,|X| - \text{rank}B$). But this contradicts Accessibility, proving that $\text{rank}B = |X| - 1$. This completes the proof.

B.7 Proof of Proposition 7

As argued in Section 4, the equivalence 1.a $\Leftrightarrow$ 1.b and part 2 of the statement follow directly from Proposition 5. Here I prove the equivalence 1.a $\Leftrightarrow$ 1.c. To this end, say that $C'\boldsymbol{u} = 0$ identifies $\boldsymbol{u}$ given $\boldsymbol{w}$ if

$$\forall\ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|A||X|} :\ \boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{v}),\ C'\boldsymbol{u} = C'\boldsymbol{v} \text{ and } \boldsymbol{w}(\boldsymbol{u}) = \boldsymbol{w}(\boldsymbol{v}) \ \Rightarrow\ \boldsymbol{u} = \boldsymbol{v}$$

and that $C'\boldsymbol{u} = 0$ identifies $\boldsymbol{w}$ if

$$\forall\ \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|A||X|} :\ \boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{v}) \text{ and } C'\boldsymbol{u} = C'\boldsymbol{v} \ \Rightarrow\ \boldsymbol{w}(\boldsymbol{u}) = \boldsymbol{w}(\boldsymbol{v})\ .$$

It is easy to see that $C'\boldsymbol{u} = 0$ identifies $\boldsymbol{u}$ if and only if it identifies $\boldsymbol{u}$ given $\boldsymbol{w}$ and it identifies $\boldsymbol{w}$. By Proposition 5, $C'\boldsymbol{u} = 0$ identifies $\boldsymbol{u}$ given $\boldsymbol{w}$ if and only if the system $\left[\begin{array}{cc} M_0 & C \end{array}\right]' z = 0$ admits $z = 0$ as a unique solution. This is the case if and only if the matrix $\left[\begin{array}{cc} M_0 & C \end{array}\right]'$ has full-column rank. Since C is linearly independent, this is equivalent to Condition 1.c.i. Also, notice that $C'\boldsymbol{u} = 0$ identifies $\boldsymbol{w}$ if and only if there is no $\boldsymbol{\nu} \in \mathbb{R}^{|A||X|}$ such that

$$\left[\begin{array}{cc} M_0 & C \end{array}\right]' \boldsymbol{\nu} = \left[\begin{array}{c} 1_{|M_0|} \\ 0_{|C|} \end{array}\right]$$

which is equivalent to unfeasibility of the problem

$$\sup_{\boldsymbol{\nu} \in \mathbb{R}^{|A||X|}} 0 \text{ s.t. } \left[\begin{array}{cc} M_0 & C \end{array}\right]' \boldsymbol{\nu} = \left[\begin{array}{c} 1_{|M_0|} \\ 0_{|C|} \end{array}\right].$$

By linear programming duality, this is equivalent to unboundedness of the problem

$$\inf_{\alpha \in \mathbb{R}^{|M_0|},\ \lambda \in \mathbb{R}^{|C|}} \alpha' 1_{|M_0|} \text{ s.t. } M_0\alpha + C\lambda = 0_{|A||X|}\ .$$

Notice that the above problem is unbounded if and only if there exist two vectors $\alpha \in \mathbb{R}^{|M_0|}$ and $\lambda \in \mathbb{R}^{|C|}$ such that

$$\alpha' 1_{|M_0|} \neq 0 \text{ and } M_0\alpha + C\lambda = 0_{|A||X|}\ .$$

Letting $\boldsymbol{\nu} \equiv M_0\alpha = -C\lambda$, it is easy to see that $\boldsymbol{\nu}$ satisfies Condition 1.c.ii. This completes the proof.

B.8 Proof of Proposition 9

I prove that C satisfies Condition 1.c in Proposition 7. To do this I exploit the following Lemma.

Lemma 8. *Let $\boldsymbol{\nu} \in \mathbb{R}^{|A||X|}$. Then $\boldsymbol{\nu} \in$ SpanM if and only if*

$$\sum_a \boldsymbol{\nu}(a,x) = \sum_{a',x'} \boldsymbol{\nu}(a',x')T(x|a',x') \text{ for all } x \ .$$

Proof. Let $\tilde{M}$ be the set of all $\boldsymbol{\nu} \in \mathbb{R}^{|A||X|}$ satisfying the above condition. $\tilde{M}$ is a linear space that contains Span$M$. Moreover, by the proof of Lemma 1, it follows that $\tilde{M}$ it has dimension $(|A|-1)|X|+1$. Hence $\tilde{M} =$ Span$M$ follows from the fact that, from Lemma 1, Span$M$ has also dimension $(|A|-1)|X|+1$. □

Notice that Proposition 9 can be restated as saying that $C'\boldsymbol{u}$ identifies $\boldsymbol{u}$ if and only if there exists some subset $X^* \subseteq X$ such that i) $|X^*| \geq |X|-1$, and ii) for every non-empty $Y \subseteq X^*$ there exists $y \in Y$ such that $T(Y|a,y) < 1$. Let $X^* \subseteq X$ be a non-empty subset of states. I start by showing that there exists a non-null $\alpha \in \mathbb{R}^{|X^*|}$ such that $\sum_{x\in X^*} \alpha(x)1\{a,x\} \in$ Span$M$ if and only if, for some $Y \subseteq X^*$, we have $T(Y|a,y) = 1$ for all $y \in Y$. To see this, let

$$Y = \{x \in X^* \text{ such that } \alpha_x > 0\}$$

so that $Y \neq \varnothing$ without loss of generality. If $\sum_{x\in X^*} \alpha(x)1\{a,x\} \in$ Span$M$ then, by Lemma 8, we have

$$\alpha_x = \sum_{x'\in Y} \alpha_{x'}T(x|a,x') + \sum_{x'\in X^*\setminus Y} \alpha_{x'}T(x|a,x') \text{ for all } x \in Y \ .$$

Summing these equations for all $x \in Y$ implies that

$$\sum_{x\in Y} \alpha_x = \sum_{x\in Y} \alpha_x T(Y|a,x) + \sum_{x\in X^*\setminus Y} \alpha_x T(Y|a,x) \leq \sum_{x\in Y} \alpha_x T(Y|a,x) \leq \sum_{x\in Y} \alpha_x$$

where the first inequality follows from the fact that $\alpha_x \leq 0$ for all $x \in X^* \setminus Y$. This implies that

$T(Y|a,x)=1$ for all $x\in Y$. Conversely, if there is one such y, then the set of equations in α

$$\alpha_x=\sum_{x'\in Y}\alpha_{x'}T(x|a,x') \text{ if } x\in Y$$
$$0=\sum_{x'\in Y}\alpha_{x'}T(x|a,x') \text{ if } x\notin Y$$

is satisfied if and only if

$$\alpha_x=\sum_{x'\in Y}\alpha_{x'}T(x|a,x') \text{ if } x\in Y$$

which is a set of $|Y|$ with at least one degree of freedom, since $\sum_{x\in Y}T(x|a,x')$ for all $x'\in Y$. Hence it admits a solution α such that $\sum_{x\in Y}\alpha_x\neq 0$. By Lemma 8, this implies that $\sum_{x\in X^*}\alpha(x)1\{a,x\}\in \text{Span}M$.

Now notice that, since $\dim M=(|A|-1)|X|+1$ and $\text{rank}C=|X|$, we have $\text{rank}\left[\begin{array}{cc} M_0 & C\end{array}\right]<|A||X|$ if and only if, for every $X^*\subseteq X$ such that $|X^*|\geq |X|-1$, we have $\sum_{x\in X^*}\alpha(x)1\{a,x\}\in \text{Span}M$ for some $\alpha\neq 0$. Hence $\text{rank}\left[\begin{array}{cc} M_0 & C\end{array}\right]=|A||X|$ if and only if the condition of Proposition 9 is satisfied.

By Proposition 7, it then remains to show that, under the the condition of Proposition 9, we have $\left[\begin{array}{cc} M_0 & C\end{array}\right]\lambda=0$ for some $\lambda\in\mathbb{R}^{|M_0|+|X|}$ such that $\sum_{n=1}^{|M_0|}\lambda_n\neq 0$. We have two cases. If $X^*\neq X$ then we have $T(x|a,x)=1$ for some $x\in X$, so that $1\{a,x\}\in M_0$, and we can take $\lambda=\left[\begin{array}{c} 1\{1\{a,x\}\} \\ -1\{x\}\end{array}\right]$, where $1\{1\{a,x\}\}\in\mathbb{R}^{|M_0|}$ is the vector equal to one at coordinate $1\{a,x\}\in M_0$, and zero otherwise, and $-1\{x\}\in\mathbb{R}^{|X|}$ is the vector equal to one at coordinate x, and zero otherwise. If $X^*=X$ then there exists a stationary measure $\boldsymbol{\mu}\in M$ such that $\boldsymbol{\mu}(a,x)>0$ for all $x\in X$. Then we can take λ such that $M_0[\lambda^1,...,\lambda^{|M_0|}]'=\boldsymbol{\mu}$ and $[\lambda^{|M_0|+1},...,\lambda^{|M_0|+|X|}]'=[-\boldsymbol{\mu}(a,x):\ x\in X]'$. This completes the proof.

B.9 Proof of Proposition 10

I show that Condition 1.c of Proposition 7 is satisfied. Let

$$B=\{1\{a',x'\}-1\{a,x'\}-[1\{a',x\}-1\{a,x\}]:\quad x'\neq x\}\ .$$

Following the usual notation, I treat B as a matrix with columns in $\mathbb{R}^{|A||X|}$. It follows by Lemma 8 that, for every $\alpha \in \mathbb{R}^{|X|-1}$, we have $B\alpha \in \text{Span}M$ if and only if

$$\sum_{x' \neq x} \alpha_{x'} \{T(a', x') - T(a, x') - [T(a', x) - T(a, x)]\} = 0 \ .$$

Since the set

$$\{T(a', x') - T(a, x') - [T(a', x) - T(a, x)] : \quad x' \neq x\}$$

is linearly independent, this implies $\alpha = 0_{|X|-1}$. Hence $\text{Span}B \cap \text{Span}M = \{0_{|A||X|}\}$. Since $B \subseteq C$ and $|B| = |X| - 1$, Condition 1.c.i in Proposition 7 is satisfied. To see that Condition 1.c.ii is satisfied as well notice that, since $\text{rank}B = |X| - 1$ and $\text{rank}M_0 = \dim M = (|A| - 1)|X| + 1$ by Lemma 1, we have that $\text{rank}\begin{bmatrix} M_0 & B \end{bmatrix} < |A||X|$ only if $B\alpha \in \text{Span}M_0 = \text{Span}M$ for some non-null $\alpha \in \mathbb{R}^{|X|-1}$. Hence $\text{rank}\begin{bmatrix} M_0 & B \end{bmatrix} = |A||X|$. Therefore we have that $\text{rank}\begin{bmatrix} M_0 & C \end{bmatrix} = |A||X|$ as well, so that there exist $\alpha \in \mathbb{R}^{|M_0|}$ and $\lambda \in \mathbb{R}^{|X|-1}$ such that $1\{a'', x\} = \begin{bmatrix} M_0 & B \end{bmatrix} \begin{bmatrix} \alpha \\ \lambda \end{bmatrix}$. Since all columns of M_0 sum to 1 and all columns of B sum to 0, it follows that $\sum_{\boldsymbol{\nu} \in M_0} \alpha(\boldsymbol{\nu}) = 1$. Then letting

$$\boldsymbol{\nu} \equiv M_0\alpha = -B\lambda + 1\{a'', x\}$$

it is easy to see that $\boldsymbol{\nu}$ satisfies Condition 1.c.ii. This completes the proof.

B.10 Proof of Proposition 11

The proof is based on the fact that w is smooth whenever F is regular, as stated below.

Lemma 9. *Suppose that F is regular. Then there exists a scalar $L > 0$ be such that*

$$w(v) \leq w(u) + \nabla w(u) \cdot (v - u) + L \left\| v - u \right\|_2^2 \text{ for every } u, v \in \mathbb{R}^A \ .$$

Proof. For every $a, a' \in A$ with $a \neq a'$, let $f^{a'-a}$ be a bounded density for the distribution of $\epsilon_{a'} - \epsilon_a$. Let

$\bar{f}$ be an upper bound for $f^{a'-a}$, so that, for every a, a' with $a \neq a'$, we have

$$\mathrm{P}_F[\underline{c} < \epsilon_{a'} - \epsilon_a \leq \bar{c}] = \int_{\underline{c}}^{\bar{c}} f^{a'-a}(x)\mathrm{d}x \leq \bar{f}(\bar{c} - \underline{c}) \text{ for every } \underline{c}, \bar{c} \in \mathbb{R} \text{ such that } \underline{c} < \bar{c} \ .$$

Let also

$$K = \max_{u \in \mathbb{R}^A} \|u\|_\infty \text{ s.t. } \|u\|_2 \leq 1$$

so that $\|u\|_\infty \leq K \|u\|_2$ for all $u \in \mathbb{R}^A$, and define

$$L \equiv 2K\bar{f}\,|A|\,(|A|-1) \ .$$

For every $u \in \mathbb{R}^A$, let $\sigma(u) \in \Delta A$ be the vector of static choice probabilities rationalized by u. First I show that, for every $u, v \in \mathbb{R}^A$, we have $\|\sigma(v) - \sigma(u)\|_1 \leq \frac{L}{K} \|v - u\|_\infty$. To see this, fix an action a, and for every $a' \neq a$ denote

$$\Delta u(a') = u(a) - u(a'),\ \Delta v(a') = v(a) - v(a'),\ \bar{\Delta}(a') = \max\{\Delta u(a'), \Delta v(a')\},\ \underline{\Delta}(a') = \min\{\Delta u(a'), \Delta v(a')\}$$

We have

$$\begin{aligned}
&|\sigma(v)(a) - \sigma(u)(a)| = \left|\mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \Delta v(a') \ \forall\ a' \neq a] - \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \Delta u(a') \ \forall\ a' \neq a]\right| \\
\leq\ & \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \bar{\Delta}(a') \ \forall\ a' \neq a] - \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \underline{\Delta}(a') \ \forall\ a' \neq a] \\
=\ & \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \underline{\Delta}(a') \ \forall\ a' \neq a] + \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \bar{\Delta}(a') \ \forall\ a' \neq a \text{ and } \exists a' \neq a \text{ s.t. } \epsilon_{a'} - \epsilon_a > \underline{\Delta}(a')] \\
&- \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \underline{\Delta}(a') \ \forall\ a' \neq a] \\
=\ & \mathrm{P}_F[\epsilon_{a'} - \epsilon_a \leq \bar{\Delta}(a') \ \forall\ a' \neq a \text{ and } \exists a' \neq a \text{ s.t. } \epsilon_{a'} - \epsilon_a > \underline{\Delta}(a')] \leq \mathrm{P}_F[\exists\ a' \neq a \text{ s.t. } \underline{\Delta}(a') < \epsilon_{a'} - \epsilon_a \leq \bar{\Delta}(a')] \\
\leq\ & \sum_{a' \neq a} \mathrm{P}_F[\underline{\Delta}(a') < \epsilon_{a'} - \epsilon_a \leq \bar{\Delta}(a')] \leq \bar{f} \sum_{a' \neq a} [\bar{\Delta}(a') - \underline{\Delta}(a')] = \bar{f} \sum_{a' \neq a} \left|v(a) - u(a) - [v(a') - u(a')]\right| \\
\leq\ & \bar{f} \sum_{a' \neq a} [|v(a') - u(a')| + |v(a) - u(a)|] \leq 2\bar{f}(|A| - 1) \max_a |v(a) - u(a)| = 2\bar{f}(|A| - 1) \|v - u\|_\infty
\end{aligned}$$

Hence we have $\|\sigma(v) - \sigma(u)\|_1 \leq 2\bar{f}\,|A|\,(|A|-1)\,\|v - u\|_\infty = \frac{L}{K}\|v - u\|_\infty$ as we wanted to show. Now,

since w is continuously differentiable by Theorem 2, by the Fundamental Theorem of Calculus we have

$$w(v) - w(u) = \int_0^1 \nabla w(u + \alpha(v-u)) \cdot (v-u) \mathrm{d}\alpha \text{ for every } u, v \in \mathbb{R}^A \ .$$

This implies that, for every $u, v \in \mathbb{R}^A$

$$\begin{aligned} w(v) - w(u) - \nabla w(u) \cdot (v-u) &= \int_0^1 [\nabla w(u + \alpha(v-u)) - \nabla w(u)] \cdot (v-u) \mathrm{d}\alpha \\ = \int_0^1 [\sigma(u + \alpha(v-u)) - \sigma(u)] \cdot (v-u) \mathrm{d}\alpha &\leq \int_0^1 \|\sigma(u + \alpha(v-u)) - \sigma(u)\|_1 \|v-u\|_\infty \mathrm{d}\alpha \\ \leq \frac{L}{K} \int_0^1 \|u + \alpha(v-u) - u\|_\infty \|v-u\|_\infty \mathrm{d}\alpha &= \frac{L}{K} \|v-u\|_\infty^2 \leq L \|v-u\|_2^2 \end{aligned}$$

which completes the proof. □

Given this Lemma, the main statement follows by standard reasoning. To see this, let L be as in previous Lemma, and define $\bar{\gamma} \equiv \frac{1}{2L}$. In what follows, fix $x \in X$, and denote $u^* \equiv \nabla w^*(\boldsymbol{\mu})(x)$ and $\sigma \equiv \boldsymbol{\sigma}(x)$. Consider the problem

$$\max_{u \in \mathbb{R}^A} [\sigma \cdot u - w(u)].$$

Then $u = u^*$ if and only if u solves the above problem and $w(u) = 0$. For every $n \geq 1$ we have $w(u^n) = w(u^{n+1}) = 0$ by construction. Therefore, if $\gamma \leq \bar{\gamma}$, we have

$$\begin{aligned} \sigma \cdot u^{n+1} &= \sigma \cdot u^n + \gamma \sigma \cdot [\sigma - \sigma^n] - w(u^n + \gamma[\sigma - \sigma^n]) \\ &\geq \sigma \cdot u^n + \gamma \sigma \cdot [\sigma - \sigma^n] - w(u^n) - \gamma \nabla w(u^n) \cdot [\sigma - \sigma^n] - L\gamma^2 \|\sigma - \sigma^n\|_2^2 \\ &= \sigma \cdot u^n + \gamma[\sigma - \sigma^n] \cdot [\sigma - \sigma^n] - L\gamma^2 \|\sigma - \sigma^n\|_2^2 \\ &= \sigma \cdot u^n + \gamma[1 - L\gamma] \|\sigma - \sigma^n\|_2^2 \geq \sigma \cdot u^n + \frac{\gamma}{2} \|\sigma - \sigma^n\|_2^2 \end{aligned}$$

which is what we wanted to show.

B.11 Proof of Proposition 12

For every $\boldsymbol{\mu} \in M$, let $\underline{\boldsymbol{\mu}}_X = \min_{x \in X} \boldsymbol{\mu}_X(x)$ be the lowest value of $\boldsymbol{\mu}_X$. For every $\epsilon > 0$, define

$$M(\epsilon) \equiv \{\boldsymbol{\mu} \in M : \ \underline{\boldsymbol{\mu}}_X \geq 0\}$$

be the set of stationary long-run frequencies $\boldsymbol{\mu}$ such that $\boldsymbol{\mu}_X$ is bounded below by ϵ, and

$$U(\epsilon) \equiv \{\boldsymbol{u} \in \mathbb{R}^{|A||X|} : \ \boldsymbol{\mu}(\boldsymbol{u}) \in M(\epsilon)\}$$

be the set of choice-specific payoff vectors rationalizing such state-action frequencies. The proof is based on the following Lemmata.

Lemma 10. *For every $\epsilon > 0$ there exists a constant $L(\epsilon) > 0$ such that, for every $\boldsymbol{u}, \boldsymbol{v} \in U(\epsilon)$, we have*

$$\boldsymbol{w}(\boldsymbol{v}) \leq \boldsymbol{w}(\boldsymbol{u}) + \boldsymbol{\mu}(\boldsymbol{u}) \cdot [\boldsymbol{v} - \boldsymbol{u}] + L(\epsilon) \left\| \boldsymbol{v} - \boldsymbol{u} \right\|_2^2 \ .$$

Proof. Fix $\epsilon > 0$. By Lemma 4, the equivalence of norms on $\mathbb{R}^A$, and Lemma 9, it follows that there exists a scalar $K > 0$ such that

$$w^*(\alpha\sigma + (1-\alpha)\rho) \leq \alpha w^*(\sigma) + (1-\alpha) w^*(\rho) - \alpha(1-\alpha) K \left\| \sigma - \rho \right\|_1^2 \text{ for all } \sigma, \rho \in \Delta A.$$

By the duality between smoothness and strong convexity again, in order to prove the statement it suffices to show that there exists a constant $K(\epsilon) > 0$ such that, for every $\boldsymbol{\mu}, \boldsymbol{\nu} \in M(\epsilon)$ and $\alpha \in (0,1)$, we have

$$\boldsymbol{w}^*(\alpha\boldsymbol{\mu} + (1-\alpha)\boldsymbol{\nu}) \leq \alpha \boldsymbol{w}^*(\boldsymbol{\mu}) + (1-\alpha)\boldsymbol{w}^*(\boldsymbol{\nu}) - \alpha(1-\alpha) K(\epsilon) \left\| \boldsymbol{\mu} - \boldsymbol{\nu} \right\|_1^2 \ .$$

In order to show this, I first show that there exists $k(\epsilon) > 0$ such that, for every $\boldsymbol{\mu}, \boldsymbol{\nu} \in M(\epsilon)$, we have

$$\left\| \boldsymbol{\mu} - \boldsymbol{\nu} \right\|_1 \leq k(\epsilon) \left\| \boldsymbol{\sigma}^{\boldsymbol{\mu}} - \boldsymbol{\sigma}^{\boldsymbol{\nu}} \right\|_1 \ .$$

To see this, for every $\boldsymbol{\mu} \in M_+$, define the transition matrix $T^{\boldsymbol{\mu}} \in [0,1]^{X \times X}$ by

$$T^{\boldsymbol{\mu}}(x'|x) = \sum_a \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) T(x'|a,x) \text{ for every } x, x' \ .$$

It is well know that, for every $\boldsymbol{\mu}, \boldsymbol{\nu} \in M_+$, the distance between $\boldsymbol{\mu}_X, \boldsymbol{\nu}_X$ can be bounded by a multiple of the distance between $T^{\boldsymbol{\mu}}$ and $T^{\boldsymbol{\nu}}$. A survey of perturbation bounds for stationary distributions of Markov chains is presented by Cho and Meyer [2001]. In particular, they show that there exists a continuous function $k(\cdot): \ M_+ \to \mathbb{R}_+$ such that for every $\boldsymbol{\mu}, \boldsymbol{\nu} \in M_+$, we have

$$\sum_x |\boldsymbol{\mu}_X(x) - \boldsymbol{\nu}_X(x)| \leq k(\boldsymbol{\mu}) \max_x \sum_{x'} \left| T^{\boldsymbol{\mu}}(x'|x) - T^{\boldsymbol{\nu}}(x'|x) \right| \ .$$

Hence, by Weierstrass' extreme value theorem, we can let $k(\epsilon) = \max_{\boldsymbol{\mu} \in M(\epsilon)} k(\boldsymbol{\mu}) + 1$. From this it follows that, for every $\boldsymbol{\mu}, \boldsymbol{\nu} \in M(\epsilon)$, we have

$$\begin{aligned}
&\sum_x |\boldsymbol{\mu}_X(x) - \boldsymbol{\nu}_X(x)| \leq [k(\epsilon) - 1] \max_x \sum_{x'} \left| T^{\boldsymbol{\mu}}(x'|x) - T^{\boldsymbol{\nu}}(x'|x) \right| \\
&= [k(\epsilon) - 1] \max_x \sum_{x'} \left| \sum_a [\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) - \boldsymbol{\sigma}^{\boldsymbol{\nu}}(a,x)] T(x'|a,x) \right| \leq [k(\epsilon) - 1] \sum_{a,x} |\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) - \boldsymbol{\sigma}^{\boldsymbol{\nu}}(a,x)|
\end{aligned}$$

and therefore

$$\begin{aligned}
\|\boldsymbol{\mu} - \boldsymbol{\nu}\|_1 &= \sum_{a,x} |\boldsymbol{\mu}_X(x) \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) - \boldsymbol{\nu}_X(x) \boldsymbol{\sigma}^{\boldsymbol{\nu}}(a,x)| \\
&\leq \|\boldsymbol{\sigma}^{\boldsymbol{\mu}} - \boldsymbol{\sigma}^{\boldsymbol{\nu}}\|_1 \ + \|\boldsymbol{\mu}_X - \boldsymbol{\nu}_X\|_1 \leq k(\epsilon) \|\boldsymbol{\sigma}^{\boldsymbol{\mu}} - \boldsymbol{\sigma}^{\boldsymbol{\nu}}\|_1
\end{aligned}$$

Now, define $K(\epsilon) \equiv K \epsilon k(\epsilon)^2$, and take $\boldsymbol{\mu}, \boldsymbol{\nu} \in M(\epsilon)$ and $\alpha \in (0,1)$. For every x define $\lambda(x) \equiv \alpha \boldsymbol{\mu}_X(x) / [\alpha \boldsymbol{\mu}_X(x) + (1-\alpha) \boldsymbol{\nu}_X(x)]$, and notice that for every a, x we have

$$\boldsymbol{\sigma}^{\alpha \boldsymbol{\mu} + (1-\alpha) \boldsymbol{\nu}}(a,x) = \lambda(x) \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) + [1 - \lambda(x)] \boldsymbol{\sigma}^{\boldsymbol{\nu}}(a,x) \ .$$

Hence, by strong convexity of w^*, for every x we have

$$w^*(\lambda(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)+(1-\lambda(x))\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)) \leq \lambda(x)w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x))+(1-\lambda(x))w^*(\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x))-\lambda(x)(1-\lambda(x))K\left\|\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)-\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)\right\|_1^2$$

and therefore

$$\begin{aligned}
\boldsymbol{w}^*(\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}) &= \sum_x [\alpha\boldsymbol{\mu}(x)+(1-\alpha)\boldsymbol{\nu}(x)]w^*(\boldsymbol{\sigma}^{\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}}(x)) \\
&= \sum_x [\alpha\boldsymbol{\mu}(x)+(1-\alpha)\boldsymbol{\nu}(x)]w^*(\lambda(x)\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)+[1-\lambda(x)]\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)) \\
&\leq \alpha\boldsymbol{w}^*(\boldsymbol{\mu})+(1-\alpha)\boldsymbol{w}^*(\boldsymbol{\nu})-\alpha(1-\alpha)K\sum_x \frac{\boldsymbol{\mu}(x)\boldsymbol{\nu}(x)}{\alpha\boldsymbol{\mu}(x)+(1-\alpha)\boldsymbol{\nu}(x)}\left\|\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)-\boldsymbol{\sigma}^{\boldsymbol{\nu}}(x)\right\|_1^2 \ .
\end{aligned}$$

Since for every x we have $\boldsymbol{\mu}(x)\boldsymbol{\nu}(x)/[\alpha\boldsymbol{\mu}(x)+(1-\alpha)\boldsymbol{\nu}(x)] \geq \min\{\boldsymbol{\mu}(x),\boldsymbol{\nu}(x)\} \geq \epsilon$, this implies

$$\begin{aligned}
\boldsymbol{w}^*(\alpha\boldsymbol{\mu}+(1-\alpha)\boldsymbol{\nu}) &\leq \alpha\boldsymbol{w}^*(\boldsymbol{\mu})+(1-\alpha)\boldsymbol{w}^*(\boldsymbol{\nu})-\alpha(1-\alpha)K\epsilon\left\|\boldsymbol{\sigma}^{\boldsymbol{\mu}}-\boldsymbol{\sigma}^{\boldsymbol{\nu}}\right\|_1^2 \\
&\leq \alpha\boldsymbol{w}^*(\boldsymbol{\mu})+(1-\alpha)\boldsymbol{w}^*(\boldsymbol{\nu})-\alpha(1-\alpha)K(\epsilon)\left\|\boldsymbol{\mu}-\boldsymbol{\nu}\right\|_1^2 \ .
\end{aligned}$$

This completes the proof. □

Lemma 11. *Let $\boldsymbol{\mu} \in M_+$ and $(\boldsymbol{\mu}^n)_{n\geq 0} \subseteq M_+$ be such that $\underline{\boldsymbol{\mu}}_X^n \to 0$ as $n \to \infty$. Then $\boldsymbol{\mu}\cdot\nabla\boldsymbol{w}^*(\boldsymbol{\mu}^n) \to -\infty$ as $n \to \infty$.*

Proof. First, note that we must have $\min_{a,x}\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n) \to 0$ as $n \to \infty$. Let $\boldsymbol{u} = \nabla\boldsymbol{w}^*(\boldsymbol{\mu})$ and, for every $n \geq 0$, let $a_n, x_n = \arg\min_{a,x}\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)$ and $a'_n = \arg\max_a \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)$, so that $\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n) \to 0$ and $\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a'_n,x_n) \geq \frac{1}{|A|}$. Notice that, by Assumption 2, this implies that

$$[\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a_n,x_n)-\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)][\nabla\boldsymbol{w}^*(\boldsymbol{\mu}^n)(a_n,x_n)-\nabla\boldsymbol{w}^*(\boldsymbol{\mu}^n)(a'_n,x_n)] \to -\infty \text{ as } n \to \infty \ .$$

For every $n \geq 0$, define $\boldsymbol{\rho}^n \in (\mathbb{R}^A)^X$ by

$$\boldsymbol{\rho}^n(a,x) = \begin{cases} \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a,x) & \text{if } a,x \neq a_n, x_n \text{ and } a,x \neq a'_n, x_n \\ \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n) & \text{if } a,x = a_n, x_n \\ \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a'_n,x_n) + \boldsymbol{\sigma}^{\boldsymbol{\mu}}(a_n,x_n) - \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n) & \text{if } a,x = a'_n, x_n \end{cases}$$

Notice that we have $\sum_a \boldsymbol{\rho}^n(a,x) = 1$ for all x and, since $\boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n) \to 0$, $\boldsymbol{\rho}^n(a,x) \geq 0$ for all a,x for n high enough. Therefore, for n high enough, $\boldsymbol{\rho}^n \in (\Delta A)^X$ and

$$\begin{aligned} \boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n) &= \sum_{x \neq x_n} \boldsymbol{\mu}_X(x) \boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x) + \boldsymbol{\mu}_X(x_n) \boldsymbol{\sigma}^{\boldsymbol{\mu}}(x_n) \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x_n) \\ &\leq \sum_{x \neq x_n} \boldsymbol{\mu}_X(x) w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) + \boldsymbol{\mu}_X(x_n) \boldsymbol{\sigma}^{\boldsymbol{\mu}}(x_n) \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x_n) \\ &= \sum_{x \neq x_n} \boldsymbol{\mu}_X(x) w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) + \boldsymbol{\mu}_X(x_n) \boldsymbol{\rho}^n(x_n) \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x_n) \\ &\quad + [\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a_n,x_n) - \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)][\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a_n,x_n) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a'_n,x_n)] \\ &\leq \sum_{x \neq x_n} \boldsymbol{\mu}_X(x) w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) + \boldsymbol{\mu}_X(x_n) w^*(\boldsymbol{\rho}^n(x_n)) \\ &\quad + [\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a_n,x_n) - \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)][\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a_n,x_n) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a'_n,x_n)] \end{aligned}$$

Here, the first inequality follows from the fact that $\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x) \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x) \leq w^*(\boldsymbol{\sigma}^{\boldsymbol{\mu}}(x)) + w(\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x))$ by Theorem 2 and $w(\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(x)) = 0$ by Lemma 2, and the second inequality follows from the same argument. Notice that by Proposition 4 the function w^* is bounded above by $-\mathrm{E}_F \min_a \epsilon(a)$, hence this implies

$$\boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n) \leq -\mathrm{E}_F \min_a \epsilon(a) + [\boldsymbol{\sigma}^{\boldsymbol{\mu}}(a_n,x_n) - \boldsymbol{\sigma}^{\boldsymbol{\mu}^n}(a_n,x_n)][\nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a_n,x_n) - \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n)(a'_n,x_n)]$$

Since the second term goes to $-\infty$ this implies that $\boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}^n) \to -\infty$ as we wanted to show. □

To prove the main statement, fix an arbitrary constant $b > 0$ and define

$$\epsilon = \min_u \min_x \boldsymbol{\mu}(\boldsymbol{u})_X(x) \text{ s.t. } \boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u}) \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^0 - \boldsymbol{w}(\boldsymbol{u}^0) - 2 - \frac{b^2}{2}$$

Note that $\epsilon > 0$, since if $\min_x \boldsymbol{\mu}(\boldsymbol{u})_X(x) \to 0$ then $\boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u}) = \boldsymbol{\mu} \cdot \nabla \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u})) \to -\infty$ by Lemma 11. Take $L(\epsilon)$ as in Lemma 10 and let

$$\gamma = \min\{\frac{b}{2}, \frac{1}{2L(\epsilon)}\}$$

I show that, at any step $n \geq 0$ such that $\boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^0 - \boldsymbol{w}(\boldsymbol{u}^0)$, we have

$$\boldsymbol{\mu} \cdot \boldsymbol{u}^{n+1} - \boldsymbol{w}(\boldsymbol{u}^{n+1}) \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) + \frac{\gamma}{2} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^n) \right\|_2^2$$

By induction this then will then show that the above inequality holds for every $n \geq 0$. Suppose that at step n we have $\boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^0 - \boldsymbol{w}(\boldsymbol{u}^0)$. Since $\left\| \boldsymbol{u}^{n+1} - \boldsymbol{u}^n \right\|_2 = \gamma \left\| \boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}) \right\|_2 \leq 2\gamma \leq b$ and $\boldsymbol{\mu}(\boldsymbol{u}^{n+1}) \cdot [\boldsymbol{u}^n - \boldsymbol{u}^{n+1}] \leq \boldsymbol{w}(\boldsymbol{u}^n) - \boldsymbol{w}(\boldsymbol{u}^{n+1})$ by Theorem 2, we have

$$\begin{aligned}
&\boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) - [\boldsymbol{\mu} \cdot \boldsymbol{u}^{n+1} - \boldsymbol{w}(\boldsymbol{u}^{n+1})] = \boldsymbol{\mu} \cdot [\boldsymbol{u}^n - \boldsymbol{u}^{n+1}] - \boldsymbol{w}(\boldsymbol{u}^n) + \boldsymbol{w}(\boldsymbol{u}^{n+1}) \\
&\leq [\boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^{n+1})] \cdot [\boldsymbol{u}^n - \boldsymbol{u}^{n+1}] \leq \frac{\left\| \boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^{n+1}) \right\|_2^2 + \left\| \boldsymbol{u}^n - \boldsymbol{u}^{n+1} \right\|_2^2}{2} \leq 2 + \frac{b^2}{2}
\end{aligned}$$

hence

$$\boldsymbol{\mu} \cdot \boldsymbol{u}^{n+1} - \boldsymbol{w}(\boldsymbol{u}^{n+1}) \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) - 2 - \frac{b^2}{2} \geq \boldsymbol{\mu} \cdot \boldsymbol{u}^0 - \boldsymbol{w}(\boldsymbol{u}^0) - 2 - \frac{b^2}{2} \ .$$

This implies that $\min_x \boldsymbol{\mu}(\boldsymbol{u}^{n+1})_X(x) \geq \epsilon$, hence

$$\boldsymbol{w}(\boldsymbol{u}^{n+1}) \leq \boldsymbol{w}(\boldsymbol{u}^n) + \boldsymbol{\mu}(\boldsymbol{u}^n) \cdot [\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] + L(\epsilon) \left\| \boldsymbol{u}^{n+1} - \boldsymbol{u}^n \right\|_2^2$$

and finally

$$\begin{aligned}
\boldsymbol{\mu} \cdot \boldsymbol{u}^{n+1} - \boldsymbol{w}(\boldsymbol{u}^{n+1}) &\geq \boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) + [\boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^n)] \cdot [\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] - L(\epsilon) \left\| \boldsymbol{u}^{n+1} - \boldsymbol{u}^n \right\|_2^2 \\
&= \boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) + \gamma[1 - L(\epsilon)\gamma] \left\| \boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^n) \right\|_2^2 \\
&\geq \boldsymbol{\mu} \cdot \boldsymbol{u}^n - \boldsymbol{w}(\boldsymbol{u}^n) + \frac{\gamma}{2} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}(\boldsymbol{u}^n) \right\|_2^2
\end{aligned}$$

completing the proof.

B.12 Proof of Proposition 13

The proof is based on the following Lemma.

Lemma 12. *Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|A||X|}$ be such that there is no $k \in \mathbb{R}$ such that $M_0'\boldsymbol{u} = M_0'\boldsymbol{v} + k$. Then, for every scalar $\lambda \in (0,1)$, we have*

$$\boldsymbol{w}(\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}) < \lambda \boldsymbol{w}(\boldsymbol{u}) + (1-\lambda)\boldsymbol{w}(\boldsymbol{v}) \ .$$

Proof. Notice that weak inequality holds by convexity of $\boldsymbol{w}$. Suppose that $\boldsymbol{w}(\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}) = \lambda \boldsymbol{w}(\boldsymbol{u}) + (1-\lambda)\boldsymbol{w}(\boldsymbol{v})$ for some $\lambda \in (0,1)$, and denote $\boldsymbol{u}^\lambda \equiv \lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}$. By Theorem 2 we have

$$\lambda \boldsymbol{w}(\boldsymbol{u}) + (1-\lambda)\boldsymbol{w}(\boldsymbol{v}) = \lambda[\boldsymbol{\mu}(\boldsymbol{u}) \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}))] + (1-\lambda)[\boldsymbol{\mu}(\boldsymbol{v}) \cdot \boldsymbol{v} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{v}))]$$
$$= \boldsymbol{\mu}(\boldsymbol{u}^\lambda) \cdot [\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}] - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}^\lambda)) = \lambda[\boldsymbol{\mu}(\boldsymbol{u}^\lambda) \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}^\lambda))] + (1-\lambda)[\boldsymbol{\mu}(\boldsymbol{u}^\lambda) \cdot \boldsymbol{v} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}^\lambda)]$$

This implies that if $\boldsymbol{w}(\boldsymbol{u}) > \boldsymbol{\mu}(\boldsymbol{u}^\lambda) \cdot \boldsymbol{u} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}^\lambda))$ then $\boldsymbol{w}(\boldsymbol{v}) < \boldsymbol{\mu}(\boldsymbol{u}^\lambda) \cdot \boldsymbol{v} - \boldsymbol{w}^*(\boldsymbol{\mu}(\boldsymbol{u}^\lambda)$, a contradiction. Hence $\boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{u}^\lambda)$ By Theorem 2. Similarly, we have $\boldsymbol{\mu}(\boldsymbol{v}) = \boldsymbol{\mu}(\boldsymbol{u}^\lambda)$. This implies $\boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}(\boldsymbol{v})$, hence by Proposition 5 there exists $k \in \mathbb{R}$ such that $M_0'u = M_0'v + k$. □

To prove the main statement, notice that the if direction is trivial, since if $M_0'\boldsymbol{u} = M_0'\boldsymbol{v} + k$ then $M_0'(\boldsymbol{u} + \boldsymbol{\eta}) = M_0'(\boldsymbol{v} + \boldsymbol{\eta}) + k$, hence $\boldsymbol{\mu}(\boldsymbol{u} + \boldsymbol{\eta}) = \boldsymbol{\mu}(\boldsymbol{v} + \boldsymbol{\eta})$, for every $\boldsymbol{\eta}$ in the support of G, which implies $\boldsymbol{\mu}(\boldsymbol{u}; G) = \boldsymbol{\mu}(\boldsymbol{v}; G)$. For the converse, suppose that there is no $k \in \mathbb{R}$ such that $M_0'\boldsymbol{u} = M_0'\boldsymbol{v} + k$. Then, for every $\boldsymbol{\eta}$ in the support of G, there is no $k \in \mathbb{R}$ such that $M_0'(\boldsymbol{u} + \boldsymbol{\eta}) = M_0'(\boldsymbol{v} + \boldsymbol{\eta}) + k$, hence $\boldsymbol{w}(\lambda(\boldsymbol{u} + \boldsymbol{\eta}) + (1-\lambda)(\boldsymbol{v} + \boldsymbol{\eta})) < \lambda \boldsymbol{w}(\boldsymbol{u} + \boldsymbol{\eta}) + (1-\lambda)\boldsymbol{w}(\boldsymbol{v} + \boldsymbol{\eta})$ for every $\lambda \in (0,1)$ by previous Lemma. This implies that $\boldsymbol{w}(\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}; G) < \lambda \boldsymbol{w}(\boldsymbol{u}; G) + (1-\lambda)\boldsymbol{w}(\boldsymbol{v}; G)$. Suppose by contradiction that $\boldsymbol{\mu}(\boldsymbol{u}; G) = \boldsymbol{\mu}(\boldsymbol{v}; G) \equiv \boldsymbol{\mu}$, and take any $\lambda \in (0,1)$. We have

$$\boldsymbol{\mu} \cdot [\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}] - \boldsymbol{w}(\lambda \boldsymbol{u} + (1-\lambda)\boldsymbol{v}; G) > \lambda[\boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u}; G)] + (1-\lambda)[\boldsymbol{\mu} \cdot \boldsymbol{v} - \boldsymbol{w}(\boldsymbol{v}; G)] = \max_{\boldsymbol{u}' \in \mathbb{R}^{|A||X|}} [\boldsymbol{\mu} \cdot \boldsymbol{u}' - \boldsymbol{w}(\boldsymbol{u}'; G)]$$

a contradiction. Hence $\boldsymbol{\mu}(\boldsymbol{u}; G) \neq \boldsymbol{\mu}(\boldsymbol{v}; G)$, completing the proof.

B.13 Proof of Proposition 14

The only thing to show that, for every $\boldsymbol{\mu} \in M_+$ there exists $\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}$ such that $\boldsymbol{\mu}(\bar{\boldsymbol{u}}; G) = \boldsymbol{\mu}$. To prove this, define the convex conjugate $\boldsymbol{w}^*(\cdot; G)$ of $\boldsymbol{w}(\cdot; G)$:

$$\boldsymbol{w}^*(\boldsymbol{\mu}; G) = \sup_{\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}} [\boldsymbol{\mu} \cdot \bar{\boldsymbol{u}} - \boldsymbol{w}(\bar{\boldsymbol{u}}; G)] \text{ for all } \boldsymbol{\mu} \in M \ .$$

Recall that $\bar{\boldsymbol{u}}$ rationalizes $\boldsymbol{\mu}$ if and only if it solves the above problem, which is equivalent to $\bar{\boldsymbol{u}} \in \nabla \boldsymbol{w}^*(\boldsymbol{\mu}; G)$. Then it is sufficient to show that $\nabla \boldsymbol{w}^*(\boldsymbol{\mu}; G)$ is non-empty for every $\boldsymbol{\mu} \in M_+$. By Rockafellar [1970, Theorem 23.4], this would follow if $\boldsymbol{w}^*(\boldsymbol{\mu}; G) < +\infty$ for all $\boldsymbol{\mu} \in M_+$. To see this note that for every $\bar{\boldsymbol{u}}$ we have

$$\boldsymbol{w}(\bar{\boldsymbol{u}}; G) = \int \boldsymbol{w}(\bar{\boldsymbol{u}} + \boldsymbol{\eta}) \mathrm{d}G(\boldsymbol{\eta}) \geq \boldsymbol{w}(\int \bar{\boldsymbol{u}} + \boldsymbol{\eta} \mathrm{d}G(\boldsymbol{\eta})) = \boldsymbol{w}(\bar{\boldsymbol{u}})$$

by Jensen's inequality. Hence, for all $\boldsymbol{\mu} \in M_+$, taking $\boldsymbol{u}$ such that $\boldsymbol{\mu}(\boldsymbol{u}) = \boldsymbol{\mu}$, we have

$$\sup_{\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}} [\boldsymbol{\mu} \cdot \bar{\boldsymbol{u}} - \boldsymbol{w}(\bar{\boldsymbol{u}}; G)] \leq \sup_{\bar{\boldsymbol{u}} \in \mathbb{R}^{|A||X|}} [\boldsymbol{\mu} \cdot \bar{\boldsymbol{u}} - \boldsymbol{w}(\bar{\boldsymbol{u}})] = \boldsymbol{\mu} \cdot \boldsymbol{u} - \boldsymbol{w}(\boldsymbol{u}) < +\infty \ .$$

This completes the proof.