ONLINE APPENDIX TO: "DISCRETE-CONTINUOUS HIGH-DIMENSIONAL DYNAMIC OPTIMIZATION"

LORETTI I. DOBRESCU* AND AKSHAY SHANKER†

1. Preliminaries

Throughout the mathematical appendix, we will use ***bold italic*** letters to denote variables that are vectors. Abstract vector spaces will be denoted by bold italic $\boldsymbol{C}$apital letters.

Recall that for a real topological vector space (hereby referred to as a 'vector space') $\boldsymbol{X}$, $\boldsymbol{X}^\star$ denotes the topological dual, or the space of all continuous linear functions on $\boldsymbol{X}$. For an element $\boldsymbol{x}$, with $\boldsymbol{x} \in \boldsymbol{X}$, and an element $\boldsymbol{x}^\star$, with $\boldsymbol{x}^\star \in X^\star$, we use $\boldsymbol{x}^\star\boldsymbol{x}$ to refer to the evaluation of $\boldsymbol{x}^\star$ at $\boldsymbol{x}$, i.e. an evaluation of the mapping $\boldsymbol{x} \mapsto \langle \boldsymbol{x}^\star, \boldsymbol{x}\rangle$.[1] Moreover, let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be vector spaces. For a function f, with $\mathrm{f}\colon \boldsymbol{X} \to \boldsymbol{Y}$, and for $\boldsymbol{y}^\star \in \boldsymbol{Y}^\star$, we will use $\boldsymbol{y}^\star\mathrm{f}$ to denote the mapping $\boldsymbol{x} \mapsto \langle \mathrm{f}(\boldsymbol{x}), \boldsymbol{y}^\star\rangle$ and $\boldsymbol{y}^\star f(\boldsymbol{x})$ to denote the evaluation of the mapping.

We will use $\mathbf{0}$ to refer to the null vector in a vector space and use $\mathbb{R}_{+\infty}(\mathbb{R}_{-\infty})$ to denote the space $\mathbb{R} \cup \{+\infty\}(\mathbb{R} \cup \{-\infty\})$ and $\bar{\mathbb{R}}$ to denote the space $\mathbb{R} \cup \{+\infty, -\infty\}$. The effective domain of a functional f, $\operatorname{dom}\mathrm{f}$, will be all $\boldsymbol{x} \in \boldsymbol{X}$ such that $\mathrm{f}(\boldsymbol{x}) \in \mathbb{R}$.

Definition 1. ***(Concave function)*** *Let $\boldsymbol{X}$ be a vector space and let $\boldsymbol{Y}$ be a vector space specified with a positive cone P. A mapping $\mathrm{f}\colon \boldsymbol{X} \to \boldsymbol{Y}$ is concave if $\operatorname{dom}\mathrm{f}$ is convex and for all $\eta \in (0,1)$:*

$$\mathrm{f}(\eta\boldsymbol{x} + (1-\eta)\boldsymbol{y}) \geq \eta\mathrm{f}(\boldsymbol{x}) + (1-\eta)\mathrm{f}(\boldsymbol{y}), \qquad \boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom}\mathrm{f}$$

When $\boldsymbol{X}$ is a Hilbert space of square integrable random variables, for a real valued function f, we define an integral operator on $\boldsymbol{X}$ as follows.[2]

Definition 2. ***(Integral operator)*** *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, let $\boldsymbol{X} = L^2(\Omega, \mathbb{R}^N, \Sigma)$ and suppose $\mathrm{f}\colon \mathbb{R}^N \times \mathbb{R}^K \to \bar{\mathbb{R}}$. For fixed $\boldsymbol{y} \in L^2(\Omega, \mathbb{R}^K, \Sigma)$, the integral operator with kernal f is the operator $\mathrm{I}_\mathrm{f}\colon \boldsymbol{X} \to \bar{\mathbb{R}}$ defined by the evaluation $\mathrm{I}_\mathrm{f}(\boldsymbol{x}) = \int \mathrm{f}(\boldsymbol{x}(\omega), \boldsymbol{y}(\omega))\mathbb{P}(d\omega)$.*

Definition 3. ***(Sup-differential)*** *Let $\boldsymbol{X}$ be a normed space and let $\mathrm{f}\colon \boldsymbol{X} \to \mathbb{R}_{-\infty}$. The sup-differential of f at $\boldsymbol{x}$, $\partial^S\mathrm{f}$, is the set such that: (i) if $\boldsymbol{x} \in \operatorname{dom}\mathrm{f}$, then $\partial^S\mathrm{f}$ is the set of $x^\star \in \boldsymbol{X}^\star$ such that for all $\boldsymbol{\nu} \in \boldsymbol{X}$:*

$$\mathrm{f}(\boldsymbol{\nu}) - \mathrm{f}(\boldsymbol{x}) \leq \langle \boldsymbol{x}^\star, \boldsymbol{\nu} - \boldsymbol{x}\rangle$$

and (ii) if $\boldsymbol{x} \notin \operatorname{dom}\mathrm{f}$, then $\partial^S\mathrm{f} = \emptyset$.

For a proof of the following, see Proposition 16.7 by Bauschke and Combettes (2010).

* School of Economics, University of New South Wales, Australia. Email: dobrescu@unsw.edu.au.

† School of Economics, University of New South Wales, Australia and ARC Center for Excellence in Population Ageing Research (CEPAR). Email: a.shanker@unsw.edu.au.

‡ The latest version of the supplementary can be found here here.

[1] We refer readers to Luenberger (1969) section 5 for an introduction to dual spaces.

[2] The integral operator here is a special case of an Urysohn operator.

Fact 1. ***(Partial and total sup-differentials)*** *Let* $\mathrm{f}\colon \prod_{i\in\mathscr{F}} \boldsymbol{X}_i \to \mathbb{R}_{-\infty}$ *where* $\mathscr{F}$*, with* $N = |\mathscr{F}|$*, indexes a finite family of normed vector spaces* $(\boldsymbol{X}_i)_{i\in\mathscr{F}}$*. Since* $\mathscr{F}$ *is finite, the norm dual of* $\prod_{i\in\mathscr{F}} X_i$ *will be* $\prod_{i\in\mathscr{F}} X_i^\star$ *with* $x^\star(x) = \langle \boldsymbol{x}, \boldsymbol{x}^\star\rangle = \sum_{i\in\mathscr{F}}\langle x_i, x_i^\star\rangle$*. Let* $\partial^S \mathrm{f}(\boldsymbol{x})$ *denote the sub-differential of* f *at* $\boldsymbol{x}$*. The evaluation mapping defines the i'th partial sup-differential as follows* $\mathrm{ev}(\partial^S \mathrm{f}, i) =\colon \partial^S \mathrm{f}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)_i$*. Fix* $\boldsymbol{x}\in\boldsymbol{X}$*, then define* $\mathrm{R}_i\colon \boldsymbol{X}_i \to \boldsymbol{X}$ *as follows, the* j'*th component of* $\mathrm{R}_i y$ *equals* $\boldsymbol{y}$ *if* $i=j$ *and* $\boldsymbol{x}_j$ *otherwise. We define the following notation:*

$$\partial^S \mathrm{f}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_N)_i\colon = \partial^S (f\circ \mathrm{R}_i)(\boldsymbol{x}_i) \tag{1}$$

Moreover, we have that $\partial^S\mathrm{f}(\boldsymbol{x}) \subset \prod_{i\in\mathscr{F}} \partial^S(\mathrm{f}\circ\mathrm{R}_i)(\boldsymbol{x}_i)$*.*

Remark 1. *Consider the setting of Proposition 1 and consider the mapping* $\boldsymbol{x}\mapsto \partial_x \mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ *on* $\boldsymbol{X}$*. The evaluation* $\partial_x\mathrm{f}(x,y)$ *will be a Euclidean vector of length* N *for all* $x\in\mathrm{dom}\,\mathrm{f}$*. As such, if* $\partial_x\mathrm{f}(\cdot,y)$ *is measurable, then* $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ *is an* $\mathbb{R}^N$ *valued random variable on* $(\Omega,\Sigma,\mathbb{P})$ *and if* $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ *has finite variance, then* $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})\in L^2(\Omega,\mathbb{R}^N,\Sigma)$*.*

Theorem 1. ***(Fermat's Theorem)****A function* f *on a normed space* $\boldsymbol{X}$ *attains its maximum at* $\boldsymbol{x}\in\boldsymbol{X}$ *if and only if* $\mathbf{0}\in\partial^S\mathrm{f}(\boldsymbol{x})$*.*

Next, consider the setting of Definition 2. Let $\mathrm{I}_{\mathrm{f}}\colon \boldsymbol{X}\to\mathbb{R}_{-\infty}$ be an integral operator with kernal f.

Proposition 1. ***(Global maxima of integral operator)*** *If (i)* $\mathrm{f}(\cdot,\boldsymbol{y})$ *is concave a.e. for fixed* $\boldsymbol{y}$*, (ii) for* $\boldsymbol{x}\in\mathrm{dom}\,\mathrm{f}$*,* $\partial\mathrm{f}_x(\boldsymbol{x},\boldsymbol{y}) = 0$ *a.e. and (iii)* $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{X}^\star$*, then* $\boldsymbol{x}$ *is a global maximum of* I_{f}*.*

Proof. Let $\boldsymbol{x}\in\mathrm{dom}\,\mathrm{f}$. By Fermat's law, $\boldsymbol{x}$ is a global maximum of I_{f} if and only if $\mathbf{0}\in\partial^S\mathrm{I}_{\mathrm{f}}(\boldsymbol{x})$. By the definition of the sup-differential, $\mathbf{0}\in\partial^S\mathrm{I}_{\mathrm{f}}(\boldsymbol{x})$ if and only if for all $\boldsymbol{\nu}\in\boldsymbol{X}$:

$$\mathrm{I}_{\mathrm{f}}(\boldsymbol{\nu}) - \mathrm{I}_{\mathrm{f}}(\boldsymbol{x}) \le \langle \mathbf{0}, \boldsymbol{\nu}-\boldsymbol{x}\rangle$$

Since $\partial_x\mathrm{f}(x,y) = 0$ a.e., we will show that (i) $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y}) = \mathbf{0}$ and that (ii) $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})\in\partial^S\mathrm{I}_{\mathrm{f}}(\boldsymbol{x})$. Showing $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y}) = \mathbf{0}$ is straightfoward since for any $\boldsymbol{\nu}^\star\in\boldsymbol{X}^\star$, we have:

$$\begin{aligned}\mathbb{P}\left\{\boldsymbol{\nu}^\star + \partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y}) \ne \boldsymbol{\nu}^\star\right\} &= \mathbb{P}\left\{\omega \,|\, \boldsymbol{\nu}^\star(\omega) + \partial_x\mathrm{f}(\boldsymbol{x})(\omega) \ne \boldsymbol{\nu}^\star(\omega)\right\} = \\ &\mathbb{P}\left\{\omega \,|\, \boldsymbol{\nu}^\star(\omega) + \partial_x\mathrm{f}(\boldsymbol{x}(\omega)) \ne \boldsymbol{\nu}^\star(\omega)\right\} = \mathbb{P}\left\{\omega \,|\, \boldsymbol{\nu}^\star(\omega) \ne \boldsymbol{\nu}^\star(\omega)\right\} = 0\end{aligned} \tag{2}$$

and thus $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y}) = \mathbf{0}$.

Next, to show $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})\in\partial^S\mathrm{I}_{\mathrm{f}}(\boldsymbol{x})$, by the definition of the integral operator,

$$\mathrm{I}_{\mathrm{f}}(\boldsymbol{\nu}) - \mathrm{I}_{\mathrm{f}}(\boldsymbol{x}) = \int \mathrm{f}(\boldsymbol{\nu}(\omega),\boldsymbol{y}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))\mathbb{P}(\mathrm{d}\omega)$$

If $\boldsymbol{\nu}\notin\mathrm{dom}\,\mathrm{f}$, then $\int \mathrm{f}(\boldsymbol{\nu}(\omega),\boldsymbol{y}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega)))\mathbb{P}(d\omega) = -\infty$ and thus $\int \mathrm{f}(\boldsymbol{\nu}(\omega),\boldsymbol{y}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))\mathbb{P}(d\omega) \le \langle\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{\nu}-\boldsymbol{x}\rangle$. Thus, $\partial_x\mathrm{f}(\boldsymbol{x},\boldsymbol{y})\in\partial^S\mathrm{I}_{\mathrm{f}}(\boldsymbol{x})$.

On the other hand, if $\boldsymbol{\nu}\in\mathrm{dom}\,\mathrm{f}$, then

$$\begin{aligned}&\int \mathrm{f}(\boldsymbol{\nu}(\omega),\boldsymbol{y}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))\mathbb{P}(\mathrm{d}\omega)\\ &\qquad\le \int \langle \partial_x\mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))),\boldsymbol{\nu}(\omega)-\boldsymbol{x}(\omega)\rangle\mathbb{P}(d\omega) = \langle\partial_x\mathrm{f}(\boldsymbol{x}),\boldsymbol{\nu}-\boldsymbol{x}\rangle\end{aligned}$$

where inequality follows since:

$$\mathbb{P}\left\{\omega \,|\, \partial_x\mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))(\boldsymbol{\nu}(\omega)-\boldsymbol{x}(\omega)) < \mathrm{f}(\boldsymbol{\nu}(\omega),\boldsymbol{y}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))\right\} = 0$$

by the concavity of f. Thus, $\partial_x \mathrm{f}(\boldsymbol{x},\boldsymbol{y}) \in \partial^S \mathrm{I}_\mathrm{f}(\boldsymbol{x})$, and since $\partial_x \mathrm{f}(\boldsymbol{x},\boldsymbol{y}) = \mathbf{0}$, we have $\mathbf{0} \in \partial^S \mathrm{I}_\mathrm{f}(\boldsymbol{x})$. □

Proposition 2. ***(Sup-differential of integral operator with differentiable kernal)*** *Consider the setting of Definition 2. Let* $\mathrm{I}_\mathrm{f} \colon \boldsymbol{X} \to \mathbb{R}_{-\infty}$ *be an integral operator with kernel* f *a.e. in* Ω. *If* $\boldsymbol{\phi} \in \partial^S_{\boldsymbol{x}} \mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ *for* $\boldsymbol{x} \in \operatorname{dom} \mathrm{f}$ *and* $\partial_x \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))$ *is defined for* ω *a.e., then* $\boldsymbol{\phi} = \partial_x \mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ *a.e.*

Proof. By Proposition 7, $\boldsymbol{\phi}(\boldsymbol{\nu} - \boldsymbol{x}) \geq \mathrm{f}(\boldsymbol{\nu},\boldsymbol{y}) - \mathrm{f}(\boldsymbol{x},\boldsymbol{y})$ a.e. for all $\boldsymbol{\nu} \in \boldsymbol{X}$. It follows that $\boldsymbol{\phi}(\omega)$ is a sup-differential of $\mathrm{f}(\cdot,\boldsymbol{y}(\omega))$ at $\boldsymbol{x}(\omega)$ for ω a.e. in Ω. Since $\mathrm{f}(\cdot,\boldsymbol{y}(\omega))$ is concave and differentiable, its sup-differential at $\boldsymbol{x}(\omega)$ is uniqely given by $\partial_x \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))$. As such, $\boldsymbol{\phi}(\omega) = \partial_x \mathrm{f}(\boldsymbol{x}(\omega),\boldsymbol{y}(\omega))$ for ω a.e. in Ω. □

2. Proofs for Section 3

2.1. **Hilbert space problem.** This section provides a complete generalization of the problem defined in Section 2.2 to a dynamic optimization problem on a Hilbert space. Our approach will be to prove necessity and sufficiency of first order conditions for the general Hilbert space problem, and then establish that the abstract results imply the main results of 2.2.

Recall the underlying probability space on which the shocks $\{\boldsymbol{w}_t\}_{t=0}^T$ are defined is $(\Omega, \Sigma, \mathbb{P})$.[3] Let $\{\mathscr{F}_t\}_{t=0}^T$ denote the natural filtration generated by the sequence of random variables $\{w_t\}_{t=0}^T$. For an arbitrary Banach space A and sub-algebra $\bar{\Sigma} \subset \Sigma$, we use $L^2(\Omega, A, \bar{\Sigma})$ to denote the Hilbert space of A-valued $\bar{\Sigma}$- measurable and square integrable functions on Ω. Unless stated otherwise, we will equip $L^2(\Omega, A, \bar{\Sigma})$ with the weak topology.

The Hilbert space dynamic optmization problem consists of the following for each t:

(1) A state space $\boldsymbol{X}_t = \boldsymbol{W}_t \times \boldsymbol{Z}_t \times \boldsymbol{M}_t$, where:
   (i) $\boldsymbol{W}_t$ is an exogenous shock space, with $\boldsymbol{W}_t \subset L^2\left(\Omega, \mathbb{R}^{N_W}, \mathscr{F}_t\right)$.
   (ii) $\boldsymbol{Z}_t$ is a discrete state space, with $\boldsymbol{Z}_t \subset L^2\left(\Omega, \mathbb{R}^{N_Z}, \mathscr{F}_t\right)$.
   (iii) $\boldsymbol{M}_t$ is a continuous state space, with $\boldsymbol{M}_t \subset L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_t\right)$.
(2) An action space $\boldsymbol{D}_t \times \boldsymbol{Y}_t$, where:
   (i) $\boldsymbol{D}_t$ is a discrete choice space, with $\boldsymbol{D}_t \subset L^2\left(\Omega, \mathbb{N}^{N_D}, \mathscr{F}_t\right)$.
   (ii) $\boldsymbol{Y}_t$ is a continuous post-state space, with $\boldsymbol{Y}_t \subset L^2\left(\Omega, \mathbb{R}^{N_Y}, \mathscr{F}_t\right)$.
(3) A feasibility correspondence $\boldsymbol{\Gamma}_t$, with $\boldsymbol{\Gamma}_t \colon \boldsymbol{X}_t \rightrightarrows \boldsymbol{D}_t \times \boldsymbol{Y}_t$, where $\boldsymbol{d}_t, \boldsymbol{y}_t \in \boldsymbol{\Gamma}_t(\boldsymbol{x}_t)$ if and only if $\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \geq 0$ for a concave measurable function $\mathbf{g}_t$, with $\mathbf{g}_t \colon \boldsymbol{X}_t \times \boldsymbol{D}_t \times \boldsymbol{Y}_t \to L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_t\right)$.
(4) A concave measurable reward function $\mathbf{u}_t$, with $\mathbf{u}_t \colon \mathrm{Gr}\boldsymbol{\Gamma}_t \to \mathbb{R} \cup \{-\infty\}$, where we write $\mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t)$ as the evaluation of $\mathbf{u}_t$
(5) A concave measurable 'transition kernel' $\mathbf{F}_t^m$ for the continuous state, with $\mathbf{F}_t^m \colon \boldsymbol{D}_t \times \boldsymbol{Y}_t \times \boldsymbol{W}_{t+1} \to \boldsymbol{M}_{t+1}$ and a transition kernel $\mathbf{F}_t^d$, with $\mathbf{F}_t^d \colon \boldsymbol{D}_t \times \boldsymbol{W}_{t+1} \to \boldsymbol{Z}_{t+1}$ for the discrete state.

Assumption 1. *For each* t,

[3]Note the sequence $\{\boldsymbol{w}_t\}_{t=0}^T$ is a set of elements in an Hilbert space. To reduce discussion of abstract spaces in the main text, we used non-bold italic letters to refer to $\{w_t\}_{t=0}^T$ through a slight abuse of notation. In the appendix, we define $\{\boldsymbol{w}_t\}_{t=0}^T$ as a sequence in $L^2\left(\Omega, \mathbb{R}^{N_W}, \mathscr{F}_t\right)$ and distinguish between $\boldsymbol{w}_t$ and a particular realization $w_t = \boldsymbol{w}_t(\omega)$.

(1) $\mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t) = \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t)$,

(2) $\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t)$ *a.e.*

(3) $\boldsymbol{x}_{t+1} = \mathbf{F}_t(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1}) = \mathrm{F}_t(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1})$ *a.e.*

We use $\mathbf{F}_t$ to denote the tuple $(\mathrm{F}_t^w, \mathbf{F}_t^m, \mathbf{F}_t^d)$. This optimization problem resembles a 'deterministic' problem of controlling a sequence of vectors. In particular,[4]

$$(\mathbf{DV}) \qquad \mathbf{v}_0(\boldsymbol{x}_0)\colon = \max_{\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T} \sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t), \qquad \boldsymbol{x}_0 \in \boldsymbol{X}_0$$

such that for each t, $\boldsymbol{x}_t \in \boldsymbol{X}_t$, $\boldsymbol{d}_t \in \boldsymbol{D}_t$, $\boldsymbol{y}_t \in \boldsymbol{Y}_t$, $\boldsymbol{d}_t, \boldsymbol{y}_t \in \boldsymbol{\Gamma}_t(\boldsymbol{x}_t)$ and $\boldsymbol{x}_{t+1} = \mathrm{F}_t(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1})$ - i.e., such a sequence $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ of vectors is feasible.

We can also define a vector space problem given a sequence of discrete choices

$$(\mathbf{CS\text{-}DV}) \qquad \mathbf{v}_0^{\vec{\boldsymbol{d}}_0}(\boldsymbol{x}_0)\colon = \max_{\{\boldsymbol{x}_t, \boldsymbol{y}_t\}_{t=0}^T} \sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t), \qquad \boldsymbol{x}_0 \in \boldsymbol{X}_0$$

such that $\{\boldsymbol{x}_t, \boldsymbol{y}_t\}_{t=0}^T$ is feasible and $\vec{\boldsymbol{d}}_0 = \{\boldsymbol{d}_0, \dots, \boldsymbol{d}_T\}$.

We will say that a sequence $\{\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{d}_t\}_{t=0}^T$ is 'generated' by $\{\mathrm{y}_t, \mathrm{d}_t\}_{t=0}^T$ if $\boldsymbol{y}_t = \mathrm{y}_t(\boldsymbol{x}_t)$, $\boldsymbol{d}_t = \mathrm{d}_t(\boldsymbol{x}_t)$, $\boldsymbol{x}_{t+1} = \mathrm{F}_t(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_t)$ holds a.e.

The connection to problem (**DP**) is given by the following straight-forward result.

Claim 1. *If Assumption 1 holds,* $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ *solves* (**DV**) *and there exists a measurable functions* $\mathrm{y}_t\colon X_t \to Y_t$ *and* $\mathrm{d}_t\colon X_t \to D_t$ *such that* $\boldsymbol{y}_t = \mathrm{y}_t(\boldsymbol{x}_t)$ *and* $\boldsymbol{d}_t = \mathrm{d}_t(\boldsymbol{x}_t)$ *almost everywhere, then* $\{\mathrm{y}_t, \mathrm{d}_t\}_{t=0}^T$ *solves* (**DP**).

Proof. Let $\{\tilde{\mathrm{y}}_t, \tilde{\mathrm{d}}_t\}_{t=0}^T$ be any other feasible sequence of functions in the sense of (**DP**). Then, for each t, let $\tilde{\boldsymbol{y}}_t = \tilde{\mathrm{y}}_t(\boldsymbol{x}_t)$, $\tilde{\boldsymbol{d}}_t = \tilde{\mathrm{d}}_t(\boldsymbol{x}_t)$, $\tilde{\boldsymbol{x}}_{t+1} = \mathrm{F}_t(\tilde{\boldsymbol{d}}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_t)$ hold almost everywhere. Since $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ solves (**DV**), we must have:

$$\sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \geq \sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \tilde{\boldsymbol{z}}_t, \tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{d}}_t, \tilde{\boldsymbol{y}}_t) \tag{3}$$

However, we also have that:

$$\sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \tilde{\boldsymbol{z}}_t, \tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{d}}_t, \tilde{\boldsymbol{y}}_t) = \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \tilde{\boldsymbol{z}}_t, \tilde{\boldsymbol{m}}_t, \tilde{\mathrm{d}}_t(\tilde{\boldsymbol{x}}_t), \tilde{\mathrm{y}}_t(\tilde{\boldsymbol{x}}_t)) \tag{4}$$

and

$$\sum_{t=0}^{T} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \mathrm{d}_t(\boldsymbol{x}_t), \mathrm{y}_t(\boldsymbol{x}_t)) \tag{5}$$

[4]For our purposes, since w_0 is degenerate, the above value functions ((**DV**) and (**CS-DV**)) can be interpreted as real valued functions on X_0. We omitt a details of a discussion on Bellman equations for Hilbert space problems and refer readers to Hernandez-Lerma and Lasserre (2012) and also Shanker (2017).

To complete the proof, note that since $\{\tilde{\mathrm{y}}_t, \tilde{\mathrm{d}}_t\}_{t=0}^T$ was any other feasible sequence, the following holds

$$\mathbf{v}_0(\boldsymbol{x}_0) = \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \mathrm{d}_t(\boldsymbol{x}_t), \mathrm{y}_t(\boldsymbol{x}_t)) \geq \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \tilde{\boldsymbol{z}}_t, \tilde{\boldsymbol{m}}_t, \tilde{\mathrm{d}}_t(\boldsymbol{x}_t), \tilde{\mathrm{y}}_t(\boldsymbol{x}_t)) \tag{6}$$

□

Remark 2. *A similar argument verifies that* $\mathbf{v}_0^{\vec{\boldsymbol{d}}_0}(\boldsymbol{x}_0) = \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(\boldsymbol{x}_0)$.

2.2. **Sufficient Conditions.** Given a sequence of multipliers $\{\boldsymbol{\mu}_t\}_{t=0}^T$ and $\{\boldsymbol{\lambda}_t\}_{t=0}^T$ where $\boldsymbol{\lambda}_t \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_t\right)$ and $\boldsymbol{\mu}_t \in L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_t\right)$, and a sequence of discrete choices $\vec{\boldsymbol{d}}_0$, define the perturbation function $\mathbf{H}_t \colon \boldsymbol{M}_t \times \boldsymbol{Y}_t \to \bar{\mathbb{R}}$ as follows:

$$\mathbf{H}_t(\boldsymbol{m}, \boldsymbol{y}) \colon = \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}, \boldsymbol{d}_t, \boldsymbol{y}) + \boldsymbol{\lambda}_{t+1} \mathbf{F}_t^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}) + \boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}, \boldsymbol{y}) - \boldsymbol{\lambda}_t \boldsymbol{m} \tag{7}$$

The function $\mathbf{H}_t$ is a discrete time equivalent of a deterministic Hamiltonian, generalized to a vector space.[5] As we show below, the first order conditions of $\mathbf{H}_t$ are sufficient and neccesary conditions for optima to the problem (**CS-DV**). While this vector space variational approach avoids differentiating the value function using envelope conditions, deriving first order conditions of $\mathbf{H}_t$ requires us to use functional Gateux differentials on the underlying Hilbert spaces. To avoid the need to use functional derivatives when working on individual applications, our approach will be to prove general results using Gateux differentials below and show the first order conditions of the S-function are sufficient and neccesary to characterize the conditions required on $\mathbf{H}_t$. Once these general results provided, employing the S-function in specific applications becomes straightforward as it avoids both functional calculus and envelope conditions.

We start with general sufficient conditions for a feasible sequence to solve (**CS-DV**).

Proposition 3. ***(Infinite horizon sufficiency)*** *If there exists a feasible sequence* $\{\boldsymbol{m}_t, \boldsymbol{y}_t\}_{t=0}^T$ *and multipliers* $\{\boldsymbol{\lambda}_t\}_{t=0}^T$ *and* $\{\boldsymbol{\mu}_t\}_{t=0}^T$ *such that:*

(1) $\boldsymbol{m}_t \in \boldsymbol{M}_t$, $\boldsymbol{y}_t \in \boldsymbol{Y}_t$, $\boldsymbol{\lambda}_t \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_t\right)^\star$ *and* $\boldsymbol{\mu}_t \in L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_t\right)^\star$ *for each* t,

(2) $\boldsymbol{\lambda}_t \geq \boldsymbol{0}$ *and* $\boldsymbol{\mu}_t \geq \boldsymbol{0}$ *for each* t,

(3) $\boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = 0$ *for each* t,

(4) $(\boldsymbol{m}_t, \boldsymbol{y}_t) \in \arg\max_{\boldsymbol{m}, \boldsymbol{y}} \boldsymbol{H}_t(\boldsymbol{m}, \boldsymbol{y})$ *for each* t,

(5) if $T = \infty$, *then* $\lim_{t\to\infty} \boldsymbol{\lambda}_{t+1} \boldsymbol{m}_{t+1} = 0$,

then $\{\boldsymbol{m}_t, \boldsymbol{y}_t\}_{t=0}^\infty$ *solves* (**CS-DV**).

Proof. Let conditions 1.- 5. pf the proposition hold for a feasible sequence $\{\boldsymbol{m}_t, \boldsymbol{y}_t\}_{t=0}^T$ and let $\{\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_t\}_{t=0}^T$ be any feasible sequence. Consider $\Delta_{\bar{T}} \colon = \sum_{t=0}^{\bar{T}} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \sum_{t=0}^{\bar{T}} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)$ for any $\bar{T} \in \mathbb{N}$ with $\bar{T} \leq T$. If $T < \infty$, fix $\boldsymbol{\lambda}_{T+1} = \boldsymbol{0}$, $\boldsymbol{\lambda}_{T+1} \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_T\right)^\star$, and let

[5] A similar discrete time pertubation function can be found in Sorger (2015), where it is refereed to as an 'M-function'. However, the M-function by Sorger (2015) is defined on Euclidean spaces, not on a general vector space.

$\boldsymbol{m}_{T+1}, \tilde{\boldsymbol{m}}_{T+1} \in \boldsymbol{M}_T$. We have:

$$
\begin{aligned}
\Delta_{\bar{T}} \geq \sum_{t=0}^{\bar{T}} \big[& \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t) \\
& + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_t) - \boldsymbol{\lambda}_{t+1}\boldsymbol{m}_{t+1} - \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_t) + \boldsymbol{\lambda}_{t+1}\tilde{\boldsymbol{m}}_{t+1} \\
& + \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)\big]
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
= \sum_{t=0}^{T} \big[& \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t) \\
& + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_t) - \boldsymbol{\lambda}_t\boldsymbol{m}_t - \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_t) + \boldsymbol{\lambda}_t\tilde{\boldsymbol{m}}_t \\
& + \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)\big] \\
& + \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1})
\end{aligned} \tag{9}
$$

$$
= \sum_{t=0}^{T} \mathbf{H}_t(\boldsymbol{m}_t, \boldsymbol{y}_t) - \mathbf{H}_t(\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_t) + \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1}) \tag{10}
$$

$$
\geq \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1}) \tag{11}
$$

The first inequality, (8), follows from the assumption that $\boldsymbol{m}_{t+1} = \mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1}), \tilde{\boldsymbol{m}}_{t+1} = \mathbf{F}_t^m(\boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_{t+1})$, conditions 2. and 3. of the proposition and the fact that $\{\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_t\}_{t=0}^T$ is feasible. The first equality, (9), follows from re-arranging the RHS terms in the summation sign on the first line and the assumption $\boldsymbol{m}_0 = \tilde{\boldsymbol{m}}_0$. The second equality, (10), follows the definition of the function $\mathbf{H}_t$ by Equation (7). The final inequality, (11), follows by condition 4. of the proposition.

Finally, if $T < \infty$, we have $\boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1}) = 0$, and thus $\Delta_T \geq 0$. If $T = \infty$ by condition 5 of the proposition, we have:

$$
\sum_{t=0}^{\infty} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \sum_{t=0}^{\infty} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t) \geq \lim_{\bar{T}\to\infty} \boldsymbol{\lambda}_{\bar{T}}(\tilde{\boldsymbol{m}}_{\bar{T}+1} - \boldsymbol{m}_{\bar{T}+1}) \geq 0 \tag{12}
$$

whence $\Delta_\infty \geq 0$, completing the proof. □

The connection for the 'S-function' to the Hamiltonian is given by the following claim.

Claim 2. *Let* $\{\tilde{\mathrm{y}}_t, \tilde{\mathrm{d}}_t\}_{t=0}^T$ *be a measurable sequence of feasible policy functions and let* $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ *be generated by* $\{\tilde{\mathrm{y}}_t, \tilde{\mathrm{d}}_t\}_{t=0}^T$. *Moreover, let* $\boldsymbol{\lambda}_{t+1} = \Lambda_t(\boldsymbol{x}_t)$ *for each* t *and* $\boldsymbol{\mu}_t = \mathrm{\mu}_t(\boldsymbol{x}_t)$. *If Assumption 1 holds, then:*

$$
\boldsymbol{H}_t(\boldsymbol{m}, \boldsymbol{y}) = \mathrm{I}_{\mathbf{S}(\cdot, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)} \tag{13}
$$

Proof. Recall the definition of the S-function,

$$
\begin{aligned}
\mathbf{S}_t\left(m, y, \mu, \lambda \,|\, x_t, \boldsymbol{\Xi}_t\right) = {} & \mathrm{u}_t(w_t, z_t, d_t, m) + \mathbb{E}_t\Lambda_{t+1}(x_{t+1})\mathrm{F}_t^m(d_t, y, w_{t+1}) \\
& - \lambda m \\
& + \mu \mathrm{g}(w_t, z_t, m, y)
\end{aligned} \tag{$\mathbf{S}$}
$$

where, $\boldsymbol{\Xi}_t$ denotes the tuple of functions $(\mathrm{d}_t, \mathrm{y}_t, \Lambda_{t+1})$, $x_{t+1} = \mathrm{F}_t^m(d_t, \mathrm{y}_t(x_t), w_{t+1})$, $w_{t+1} = \mathrm{F}_t^w(w_t, \eta_{t+1})$, $x_{t+1} = (w_{t+1}, m_{t+1}, z_{t+1})$ and $d_t = \mathrm{d}_t(x_t)$. Moreover, we have $x_t \in X_t \subset \mathbb{R}^{N_W + N_Z + N_M}$ and similarly, $\lambda \in \mathbb{R}^{N_M}$ and $\mu \in \mathbb{R}^{N_g}$. Thus, the integral operator with kernal $\mathbf{S}_t(\cdot, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)$ can be defined by

$$\begin{aligned}\mathrm{I}_{\mathbf{S}(\cdot,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)}(\boldsymbol{m},\boldsymbol{y}) &= \int \mathbf{S}_t(\boldsymbol{m}(\omega),\boldsymbol{y}(\omega),\boldsymbol{\mu}_t(\omega),\boldsymbol{\lambda}_t(\omega)\,|\,\boldsymbol{x}_t(\omega),\Xi_t)\,\mathbb{P}(\mathrm{d}\omega)\\ &= \mathbf{u}_t(\boldsymbol{w}_t,\boldsymbol{z}_t,\boldsymbol{d}_t,\boldsymbol{m}) + \boldsymbol{\lambda}_{t+1}\mathbf{F}^m_t(\boldsymbol{d}_t,\boldsymbol{y},\boldsymbol{w}_{t+1}) - \boldsymbol{\lambda}_t\boldsymbol{m} + \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t,\boldsymbol{z}_t,\boldsymbol{m},\boldsymbol{y}) = \mathbf{H}_t(\boldsymbol{m},\boldsymbol{y})\end{aligned}$$

where the last equality follows from the definition of the S-function and by Assumption 1. □

We can now prove the first part of Proposition 1.

Proof of Proposition 1 Item (1). Let $\{\mathrm{y}_t,\mathrm{d}_t\}_{t=0}^T$ be a sequence of measurable feasible policy functions and let $\{\boldsymbol{x}_t,\boldsymbol{d}_t,\boldsymbol{y}_t\}_{t=0}^T$ be generated by $\{\mathrm{y}_t,\mathrm{d}_t\}_{t=0}^T$. Moreover, let $\boldsymbol{\lambda}_{t+1}=\Lambda_t(\boldsymbol{x}_t)$ and $\boldsymbol{\mu}_t = \upmu_t(\boldsymbol{x}_t)$ for each t. Let conditions (i)-(iv) of Proposition 1, Part (1) (the proposition to be proved) hold.

We will show that $\{\boldsymbol{m}_t,\boldsymbol{y}_t\}_{t=0}^T$ solves Problem **CS-DV** given $\vec{\boldsymbol{d}}_0=\{\boldsymbol{d}_t\}_{t=0}^T$, and thus $\{\mathrm{y}_t\}_{t=0}^T$ solves (**CS-DP**) given $\vec{\boldsymbol{d}}_0$ by Claim 1. To do so, we check that $\{\boldsymbol{m}_t,\boldsymbol{y}_t\}_{t=0}^T$ and $\{\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\}_{t=0}^T$ satisfy conditions (1)-(5) of Proposition 3. Conditions (1) - (3) follow immediately from the conditions of the proposition to be proved. Turning to condition (4), we will show that $\{\boldsymbol{m}_t,\boldsymbol{y}_t\}\in\arg\max_{\boldsymbol{m},\boldsymbol{y}}\mathbf{H}_t(\boldsymbol{m},\boldsymbol{y})$ for each t. By Claim 2, $\mathbf{H}_t(\boldsymbol{m},\boldsymbol{y}) = \mathrm{I}_{\mathbf{S}(\cdot,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)}$ and we can write,

$$\begin{aligned}\mathbf{H}_t(\boldsymbol{m},\boldsymbol{y}) &= \mathrm{I}_{\mathbf{S}(\cdot,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)}(\boldsymbol{m},\boldsymbol{y})\\ &= \int \mathbf{S}_t(\boldsymbol{m}(\omega),\boldsymbol{y}(\omega),\boldsymbol{\mu}_t(\omega),\boldsymbol{\lambda}_t(\omega)\,|\,\boldsymbol{x}_t(\omega),\Xi_t)\,\mathbb{P}(\mathrm{d}\omega)\end{aligned}$$

Thus, we will need to show that that $\partial_{(\boldsymbol{m},\boldsymbol{y})}\mathbf{S}_t(\boldsymbol{m}_t,\boldsymbol{y}_t,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)\in L^2\left(\Omega,\mathbb{R}^{N_M+N_Y},\mathscr{F}_t\right)^\star$ and that $\partial_{(\boldsymbol{m},\boldsymbol{y})}\mathbf{S}_t(\boldsymbol{m}_t,\boldsymbol{y}_t,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)=\mathbf{0}$. Square integrability is immediate from the definition of the S-function and condition (iv) of the proposition to be proved. Moreover,

$$\partial_{(\boldsymbol{m},\boldsymbol{y})}\mathbf{S}_t(\boldsymbol{m}_t,\boldsymbol{y}_t,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t) = \partial_{(m,y)}\mathbf{S}_t(\boldsymbol{m}_t,\mathrm{y}_t(\boldsymbol{x}_t),\upmu_t(\boldsymbol{x}_t),\Lambda_t(\boldsymbol{x}_t)\,|\,\boldsymbol{x}_t,\Xi_t)$$

Applying condition (i) of the proposition to be proved, $\mathbf{H}_t(\boldsymbol{m},\boldsymbol{y}) = \mathrm{I}_{\mathbf{S}(\cdot,\boldsymbol{\mu}_t,\boldsymbol{\lambda}_t\,|\,\boldsymbol{x}_t,\Xi_t)}$ then follows from Claim 2. Finally the transversality condition, condition (5) of 3, follows from condition (v) of the proposition to be proved. □

Proof of Proposition 1 Item (2). Let $\{\mathrm{y}_t,\mathrm{d}_t\}_{t=0}^T$ be a sequence of measurable feasible policy functions. Assume the conditions for part (1) hold. In the proof we will show

$$\mathrm{v}_0(x_0) = \sum_{t=0}^T \mathbb{E}_0\mathrm{u}_t(\boldsymbol{w}_t,\boldsymbol{z}_t,\boldsymbol{m}_t,\boldsymbol{d}_t,\boldsymbol{y}_t),\qquad x_0\in X_0 \tag{14}$$

holds for a stochastc recursive sequence $\{\boldsymbol{m}_t,\boldsymbol{d}_t,\boldsymbol{y}_t\}_{t=0}^T$ generated by $\{\mathrm{y}_t,\mathrm{d}_t\}_{t=0}^T$. We proceed by induction; make the inductive assumption that for $t+1<T$, $\{\mathrm{y}_t,\mathrm{d}_t\}_{j=t+1}^T$ satisfies

$$\mathrm{v}_{t+1}(x_{t+1}) = \max_{\{\mathrm{y}_j,\mathrm{d}_j\}_{j=t+1}^T}\sum_{j=t+1}^T \mathbb{E}_{t+1}\mathrm{u}_j(\boldsymbol{w}_j,\boldsymbol{z}_j,\boldsymbol{m}_j,\boldsymbol{d}_j,\boldsymbol{y}_j),\qquad x_{t+1}\in X_{t+1} \tag{15}$$

for a stochastc recursive sequence $\{\boldsymbol{m}_j,\boldsymbol{d}_j,\boldsymbol{y}_j\}_{j=t+1}^T$ adapted to the filtration $\{\mathscr{F}_j\}_{j=t+1}^T$ and generated by $\{\mathrm{y}_j,\mathrm{d}_j\}_{j=t+1}^T$.

Let $t=T$, then $\mathrm{v}_T(x_T) = \mathrm{u}_T(\boldsymbol{w}_T,\boldsymbol{z}_T,\boldsymbol{m}_T,\boldsymbol{d}_T,\boldsymbol{y}_T)$ will hold by the assumption of part (2) of the proposition to be proved. Next, fix x_t and note that given d_t, we can let $x'=\mathrm{F}_t(d_t,y_t,w_{t+1})$. From the Bellman Principle of Optimality (Hernandez-Lerma and Lasserre (2012), Section 3), we have:

$$\max_{\vec{\boldsymbol{d}}'} \mathbb{E}_t \mathrm{v}_{t+1}^{\vec{\boldsymbol{d}}'}(x') = \mathbb{E}_t \max_{\{\boldsymbol{m}_j,\boldsymbol{d}_j,\boldsymbol{y}_j\}_{j=t+1}^T} \sum_{j=t+1}^{T} \mathbb{E}_{t+1}\mathrm{u}_j(\boldsymbol{w}_j,\boldsymbol{z}_j,\boldsymbol{m}_j,\boldsymbol{d}_j,\boldsymbol{y}_j) = \mathbb{E}_t \mathrm{v}_{t+1}(x') \tag{16}$$

where maximization in the second term is over feasible stochastic resursive sequences with $\boldsymbol{x}_{t+1} = x'$.

Moreover, $\mathrm{Q}_t^{\vec{\boldsymbol{d}}'}(x_t, d_t, y_t) = \mathrm{u}_t(w_t, z_t, m_t, d_t, y_t) + \mathbb{E}_t \mathrm{v}_{t+1}^{\vec{\boldsymbol{d}}'}(x')$ for any sequence $\vec{\boldsymbol{d}}'$. However, by the assumption of part (2) of the proposition to be proved, we have:

$$\mathrm{v}_t(x_t) = \max_{d_t,\vec{\boldsymbol{d}}',y_t} \mathrm{u}_t(w_t, z_t, m_t, d_t, y_t) + \mathbb{E}_t \mathrm{v}_{t+1}(x') = \mathrm{Q}_t^{\vec{\boldsymbol{d}}_{t+1}}(x_t, \mathrm{d}_t(x_t), \mathrm{y}_t(x_t)), \qquad x_t \in X_t$$

where As such, fix $\vec{\boldsymbol{d}}_{t+1} = \{\boldsymbol{d}_{t+1}, \dots, \boldsymbol{d}_T\}$ is generated by $\{\mathrm{y}_j, \mathrm{d}_j\}_{j=t+1}^T$ and (x_t).

Now, the sequence of policy function $\{\mathrm{y}_{d_t,t}, \mathrm{y}_{t+1}, \dots, \mathrm{y}_T\}$ solves solves (**CS-DP**) given $\vec{\boldsymbol{d}}_t = \{d_t, \boldsymbol{d}_{t+1}, \dots\}$ starting at time t. As such, we have that $\mathrm{v}_t^{\vec{\boldsymbol{d}}_t}(x_t) = \mathrm{Q}_t^{\vec{\boldsymbol{d}}_t}(x_t, d_t, \mathrm{y}_t^{d_t}(x_t))$. However,

$$\mathrm{v}_t^{\vec{\boldsymbol{d}}_t}(x_t) = \max_y \mathrm{u}_t(w_t, z_t, m_t, d_t, y) + \mathbb{E}_t \mathrm{v}_{t+1}^{\vec{\boldsymbol{d}}}(x') \tag{17}$$

which implies that $\mathrm{v}_t(x_t) = \mathrm{u}_t(w_t, z_t, m_t, \mathrm{d}_t(x_t), \mathrm{y}_t(x_t)) + \mathbb{E}_t \mathrm{v}_{t+1}(x')$, where we let $x' = \mathrm{F}_t^m(\mathrm{d}_t(x_t), \mathrm{y}_t(x_t), w_{t+1})$. Noting (15), we then have:

$$\mathrm{v}_t(x_t) = \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \tag{18}$$

Thus, we have that $\mathrm{v}_t(x_t) = \sum_{t=0}^{T} \mathbb{E}_0 \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t)$ for a stochastc recursive sequence $\{\boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ generated by $\{\mathrm{y}_t, \mathrm{d}_t\}_{t=0}^T$. By the principle of induction, (15) will hold for any T. □

2.3. Necessary Conditions.

Proposition 4. *Let Assumption 1 and Assumption D.1 - D.2 hold and let* $\{\mathrm{y}_t, \mathrm{d}\}_{t=0}^T$ *be the optimal sequence. If* $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ *is generated by* $\{\mathrm{y}_t, \mathrm{d}\}_{t=0}^T$*, then there exosts* $\{\boldsymbol{\mu}_t, \boldsymbol{\lambda}_{t+1}\}_{t=0}^T$ *such that for each* t

$$\boldsymbol{\lambda}_{t+1} \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_t\right), \qquad \boldsymbol{\mu}_t \in L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_t\right) \tag{19}$$

$$\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \boldsymbol{\mu}_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}_t \boldsymbol{\lambda}_{t+1}^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = \boldsymbol{0} \tag{20}$$

$$\begin{aligned} &\partial_{\boldsymbol{m}} \mathrm{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) \\ &\qquad + \boldsymbol{\mu}_{t+1}^\top \partial_{\boldsymbol{m}} \mathrm{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) - \boldsymbol{\lambda}_{t+1}^\top = \boldsymbol{0} \end{aligned} \tag{21}$$

Proof. If $\{\boldsymbol{x}_t, \boldsymbol{d}_t, \boldsymbol{y}_t\}_{t=0}^T$ is generated by $\{\mathrm{y}_t, \mathrm{d}\}_{t=0}^T$, then $\{\boldsymbol{x}_t, \boldsymbol{y}_t\}_{t=0}^T$ solves problem (**CS-DV**) given $\vec{\boldsymbol{d}}_0 = \{\boldsymbol{d}_t\}_{t=0}^T$. As such, for each t, $\{\boldsymbol{y}_t, \boldsymbol{m}_{t+1}\}$ solve the one-shot deviation problem

$$\max_{\boldsymbol{y},\boldsymbol{m}'} \quad \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}) + \mathbf{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}', \boldsymbol{d}_{t+1}, \boldsymbol{y}') =: U_t(\boldsymbol{y}, \boldsymbol{m}_{t+1}) \tag{22}$$

subject to $\boldsymbol{y} \in \boldsymbol{Y}_t$, $\boldsymbol{m}' \in \boldsymbol{M}_{t+1}$ and the constraints

$$\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}) \geq \boldsymbol{0} \tag{23}$$

$$\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}, \boldsymbol{w}_{t+1}) - \boldsymbol{m}' = \boldsymbol{0} \tag{24}$$

$$\mathbf{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}', \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) \geq \mathbf{0} \tag{25}$$

Next, by Condition D.3, there exists $(\boldsymbol{y}, \boldsymbol{m}')$ such that $\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}) > \mathbf{0}$, $\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}, \boldsymbol{w}_{t+1}) - \boldsymbol{m}' = \mathbf{0}$ and $\mathbf{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}', \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) > \mathbf{0}$.

As such, by Theorem 1, Section 8.3 of Luenberger (1997) and applying Fermat's Theorem, there exists $\boldsymbol{\mu}_t \in L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_t\right)$, $\boldsymbol{\mu}_{t+1} \in L^2\left(\Omega, \mathbb{R}^{N_g}, \mathscr{F}_{t+1}\right)$ and $\boldsymbol{\lambda}_{t+1} \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_{t+1}\right)$ such that for each t, the following holds:

$$\begin{aligned}\mathbf{0} \in \partial^S_{(\boldsymbol{y},\boldsymbol{m}')}\big\{&U_t(\boldsymbol{y}_{t+1}, \boldsymbol{m}_{t+1}) + \boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \\ &+ \boldsymbol{\lambda}_{t+1}\left(\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1}) - \boldsymbol{m}_{t+1}\right) + \boldsymbol{\mu}_{t+1}\mathbf{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1})\big\}\end{aligned} \tag{26}$$

By Fact 1, for each t, we arrive at

$$\mathbf{0} \in \partial^S_y\left\{\mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1})\right\}$$

$$\begin{aligned}\mathbf{0} \in \partial^S_m\big\{&\mathbf{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) \\ &+ \boldsymbol{\mu}_{t+1}\mathbf{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) - \boldsymbol{\lambda}_{t+1}\boldsymbol{m}_t\big\}\end{aligned}$$

Next, note that the mapping

$$\boldsymbol{y} \mapsto \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}) + \boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}) + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}, \boldsymbol{w}_{t+1}) \tag{27}$$

is an integral operator (Definition 2) with kernal $\mathbf{S}_t(\boldsymbol{m}_t, \cdot, \boldsymbol{\mu}_t, \mathbf{0} \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)$ as follows

$$\begin{aligned}\mathrm{I}_{\mathbf{S}(\boldsymbol{m}_t, \cdot, \boldsymbol{\mu}_t, \mathbf{0} \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)}(\boldsymbol{m}, \boldsymbol{y}) &= \int \mathbf{S}_t(\boldsymbol{m}_t(\omega), \boldsymbol{y}(\omega), \boldsymbol{\mu}_t(\omega), \mathbf{0} \,|\, \boldsymbol{x}_t(\omega), \boldsymbol{\Xi}_t)\, \mathbb{P}(\mathrm{d}\omega) \\ &= \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{d}_t, \boldsymbol{m}_t, \boldsymbol{y}) + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}, \boldsymbol{w}_{t+1}) + \boldsymbol{\mu}_t \mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{y})\end{aligned}$$

Since $\partial_y \mathbf{S}_t(\boldsymbol{m}_t, \boldsymbol{y}_t, \boldsymbol{\mu}_t, \mathbf{0} \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)$ exists a.e., by Assumption D.1, and $\mathbf{S}_t(\boldsymbol{m}_t, \cdot, \boldsymbol{\mu}_t, \mathbf{0} \,|\, \boldsymbol{x}_t, \boldsymbol{\Xi}_t)$ is concave, applying Proposition 2, we have that

$$\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \boldsymbol{\mu}_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}_t \boldsymbol{\lambda}_{t+1}^\top \partial_y \mathrm{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_{t+1}) = 0$$

Similarly, we have

$$\partial_m \mathbf{S}_{t+1}(\boldsymbol{m}_{t+1}, \boldsymbol{y}_{t+1}, \boldsymbol{\mu}_{t+1}, \boldsymbol{\lambda}_{t+1} \,|\, \boldsymbol{x}_{t+1}, \boldsymbol{\Xi}_{t+1})$$

exists a.e. and $\mathbf{S}_{t+1}(\cdot, \boldsymbol{y}_t, \boldsymbol{\mu}_{t+1}, \boldsymbol{\lambda}_{t+1} \,|\, \boldsymbol{x}_{t+1}, \boldsymbol{\Xi}_{t+1})$ is concave. Applying Proposition 2, we have that

$$\begin{aligned}&\partial_m \mathrm{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) \\ &\qquad + \boldsymbol{\mu}_{t+1}^\top \partial_{\boldsymbol{m}} \mathrm{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) - \boldsymbol{\lambda}_{t+1}^\top = 0\end{aligned}$$

thereby completing the proof. □

Proof of Theorem 1. The proof proceeds to show that there exists measurable functions $\upmu_t$ and Λ_{t+1} such that for each t, the following holds:

$$\begin{aligned}&\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \\ &\qquad + \upmu_t(\boldsymbol{x}_t)^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}_t \Lambda_{t+1}(\boldsymbol{x}_t)^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = \mathbf{0}\end{aligned} \tag{28}$$

$$\begin{aligned}&\partial_{\boldsymbol{m}} \mathrm{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) \\ &\qquad + \upmu_t(\boldsymbol{x}_t)^\top \partial_{\boldsymbol{m}} \mathrm{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) - \Lambda_{t+1}(\boldsymbol{x}_t)^\top = \mathbf{0}\end{aligned} \tag{29}$$

To proceed, define $\tilde{\mu_t}\colon = \mathbb{E}\left(\mu_t \,|\, x_t\right)$ and $\tilde{\lambda}_{t+1}\colon = \mathbb{E}\left(\lambda_t \,|\, x_{t+1}\right)$ for each t. Starting with (28), we have

$$\mathbb{E}\Big[\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mu_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}_t \lambda_{t+1}^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \,\Big|\, x_t\Big] = \mathbf{0} \tag{30}$$

which gives us, using the Tower property,

$$\mathbb{E}\Big[\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mu_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \lambda_{t+1}^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \,\Big|\, x_t\Big] = \mathbf{0} \tag{31}$$

and,

$$\mathbb{E}\Big[\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mu_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}\left[\lambda_{t+1}^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) \,|\, x_{t+1}\right] \,\Big|\, x_t\Big] = \mathbf{0} \tag{32}$$

Next, by 'pulling out known factors', we have:

$$\partial_y \mathrm{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \tilde{\mu}_t^\top \partial_y \mathrm{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) + \mathbb{E}_t \tilde{\lambda}_{t+1}^\top \partial_y \mathrm{F}^m(\boldsymbol{w}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) = \mathbf{0} \tag{33}$$

and similarly for (21), we have:

$$\partial_{\boldsymbol{m}} \mathrm{u}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) + \tilde{\mu}_{t+1}^\top \partial_{\boldsymbol{m}} \mathrm{g}_{t+1}(\boldsymbol{w}_{t+1}, \boldsymbol{z}_{t+1}, \boldsymbol{m}_{t+1}, \boldsymbol{d}_{t+1}, \boldsymbol{y}_{t+1}) - \tilde{\lambda}_{t+1}^\top = \mathbf{0} \tag{34}$$

Finally, since $\tilde{\mu}_t$ is x_t measurable and $\tilde{\lambda}_{t+1}$ is x_{t+1} measurable, we can conclude that there exists measurable μ_t and Λ_{t+1} such that (20) and (21) hold.

□

2.4. **Sup-differential of the Value Function.** The following proof is presented without loss of generality with $t = 0$. The proof for $t > 0$ is analogous by reformulating the initial period the problem statements.

Proof of Corollary 1. We will show that $\partial_m \mathrm{u}_0(w_t, z_t, m_t, d_t, y_t) + \mu_0^\top \partial_m \mathrm{g}_0(w_t, z_t, m_t, d_t, y_t) \in \partial_m^S \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(w_0, z_0, m_0)$. Let $\{\boldsymbol{m}_t, \boldsymbol{y}_t\}_{t=0}^T$ be optimal given m_0 and and let $\{\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_t\}_{t=0}^T$ be optimal given any $\tilde{m}_0 \in M_0$. Consider $\Delta_{\bar{T}}\colon = \sum_{t=0}^{\bar{T}} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \sum_{t=0}^{\bar{T}} \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)$ for any $\bar{T} \in \mathbb{N}$ with $\bar{T} \leq T$. If $T < \infty$, fix $\boldsymbol{\lambda}_{T+1} = \mathbf{0}$, $\boldsymbol{\lambda}_{T+1} \in L^2\left(\Omega, \mathbb{R}^{N_M}, \mathscr{F}_T\right)$, and let $\boldsymbol{m}_{T+1}, \tilde{\boldsymbol{m}}_{T+1} \in \boldsymbol{M}_T$.

Following analogous steps as Equations (8) - (11), and noting Claim 1 and Remark 1, we have

$$\begin{aligned}
\mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(m_t) - \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(\tilde{m}_t) = \Delta_{\bar{T}} \geq & \sum_{t=0}^{\bar{T}} \big[\mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t) \\
& \quad + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_t) - \boldsymbol{\lambda}_{t+1}\boldsymbol{m}_{t+1} - \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_t) + \boldsymbol{\lambda}_{t+1}\tilde{\boldsymbol{m}}_{t+1} \\
& \quad + \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)\big] \\
= & \sum_{t=0}^{T} \big[\mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \mathbf{u}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t) \\
& \quad + \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \boldsymbol{y}_t, \boldsymbol{w}_t) - \boldsymbol{\lambda}_t\boldsymbol{m}_t - \boldsymbol{\lambda}_{t+1}\mathbf{F}_t^m(\boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t, \boldsymbol{w}_t) + \boldsymbol{\lambda}_t\tilde{\boldsymbol{m}}_t \\
& \quad + \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \boldsymbol{m}_t, \boldsymbol{d}_t, \boldsymbol{y}_t) - \boldsymbol{\mu}_t\mathbf{g}_t(\boldsymbol{w}_t, \boldsymbol{z}_t, \tilde{\boldsymbol{m}}_t, \boldsymbol{d}_t, \tilde{\boldsymbol{y}}_t)\big] \\
& + \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1}) \\
= & \sum_{t=0}^{T} \mathbf{H}_t(\boldsymbol{m}_t, \boldsymbol{y}_t) - \mathbf{H}_t(\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_t) + \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1}) \\
\geq & \ \mathbf{H}_0(\boldsymbol{m}_0, \boldsymbol{y}_0) - \mathbf{H}_0(\tilde{\boldsymbol{m}}_t, \tilde{\boldsymbol{y}}_0) + \boldsymbol{\lambda}_{T+1}(\tilde{\boldsymbol{m}}_{T+1} - \boldsymbol{m}_{T+1})
\end{aligned}$$

Multiplying both sides by -1, we get (the w_0, z_0 arguments in the future choice specific value functions are dropped for ease of notation)

$$\mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(\tilde{m}_t) - \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(m_t) \leq \mathbf{H}_0(\tilde{\boldsymbol{m}}_0, \tilde{\boldsymbol{y}}_0) - \mathbf{H}_0(\boldsymbol{m}_0, \boldsymbol{y}_0) + \boldsymbol{\lambda}_{T+1}(\boldsymbol{m}_{T+1} - \tilde{\boldsymbol{m}}_{T+1})$$

Now, since $\mathbf{H}_0$ is concave in $(m_0, \boldsymbol{y}_0)$ and since $\partial_y \mathbf{S}_0(m_0, \boldsymbol{y}_0, \boldsymbol{\mu}_0, \boldsymbol{\lambda}_0 \,|\, \Xi_t) = 0$ a.e., recalling Claim 2, we have that

$$\mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(\tilde{m}_t) - \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(m_t) \leq \partial_m \mathbf{S}_0(m_0, \boldsymbol{y}_0)(\tilde{m}_0 - m_0) + \boldsymbol{\lambda}_{T+1}(\boldsymbol{m}_{T+1} - \tilde{\boldsymbol{m}}_{T+1}) \tag{35}$$

where we have set $\boldsymbol{\lambda}_0 = 0$. If $\boldsymbol{\lambda}_{T+1} = \mathbf{0}$, then $\mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(\tilde{m}_t) - \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(m_t) \leq \partial_m \mathbf{S}_0(m_0, \boldsymbol{y}_0)(\tilde{m}_0 - m_0)$. Since $\tilde{m}_0$ is arbitrary, we have that $\partial_m \mathbf{S}_0(m_0, \boldsymbol{y}_0) \in \partial^S \mathrm{v}_0^{\vec{\boldsymbol{d}}_0}(m_0)$. On the other hand, if $T = \infty$, applying the transversality condition (Item (1.iii) of Proposition 1) delivers the required conclusion.

□

3. Proofs for Section 4.1

Proof of Claim 1. Assume $\bar{\mathrm{y}}_t$ is injective and into on a compact connected set $\bar{C}$. Call the restriction of $\bar{\mathrm{y}}_t$ to $\bar{C}$ as $\bar{\mathrm{y}}_t|_{\bar{C}}$. By Assumption I.2, $\bar{\mathrm{y}}_t|_{\bar{C}}$ is a homeomorphism on $\bar{C}$. It follows that $\bar{\mathrm{y}}_t|_{\bar{C}}(\bar{C})$ and $\bar{C}$ are a homeomorphism and must have the same dimension. Moreover, letting $U_y \subset \bar{\mathrm{y}}_t|_{\bar{C}}(\bar{C})$ such that U_y is an open neighbourhood in Y_t implies that Y_t has the same dimension as $\bar{C}$ and thus $\bar{M}_t$. □

Proof of Theorem 2. First conider that the mapping Ψ given by:[6]

$$\bar{m}, \tilde{a} \mapsto \pi_{N_Y} \nabla_a \bar{\mathbf{S}}_t(\bar{m}, \mathrm{em}_{\tilde{A}}(\tilde{a})), \qquad m \in {M_t}^{\mathrm{o}} \tag{36}$$

and note $\Psi \colon {M_t}^{\mathrm{o}} \times \left(\pi_{\tilde{A}} K_{l,t}\right)^{\mathrm{o}} \to \mathbb{R}^{N_M}$ where ${M_t}^{\mathrm{o}}$ and $\left(\pi_{\tilde{A}} K_{l,t}\right)^{\mathrm{o}}$ are both open sets in $\mathbb{R}^{N_M}$. Next, $\Psi(\cdot, \tilde{a})$ is injective for each $\tilde{a} \in \left(\pi_{\tilde{A}} K_{l,t}\right)^{\mathrm{o}}$ by (15). By the implicit function theorem (Theorem 1.1. by Kumagai (1980)), there exists an open neighbourhood $U_{0,\bar{m}}$ of $\bar{m}_t$ and an open neighbourhood

[6] We drop the ξ_t argument in the proof for easier notation.

$U_{0,\tilde{a}}$ of $\tilde{a}_t$ and a continuous function φ_0 such that for each $\tilde{a} \in U_{0,\tilde{a}}$, $\varphi_0(\tilde{a}) \in U_{0,\bar{m}}$ and $\Psi(\varphi_0(\tilde{a}), \tilde{a}) = 0$.

Next, let $\bar{\mathrm{g}}_t^{\mathrm{b}}$ be the set of binding constraint functions in the region l. Consider that by Assumption I.1, the matrix $\partial_y \bar{\mathrm{g}}_t^{\mathrm{b}}(\bar{m}, y)^\top$ is invertible for each $\bar{m}, y \in K_{l,t}$. As such, define:

$$\hat{\mu} = -\left(\partial_{\hat{y}} \bar{\mathrm{g}}_t^{\mathrm{b}}(\bar{m}, y)^\top\right)^{-1} \left[\mathbb{E}_t \Lambda_{t+1}(x')^\top \partial_{\hat{y}} \mathrm{F}_{t+1}^m \left(d_t, \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}, w_{t+1}\right) + \partial_{\hat{y}} \bar{\mathrm{u}}(\varphi(\tilde{a}), \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a})\right]$$

with $x' = \mathrm{F}(d_t, \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}, w_{t+1})$, $\hat{\mu}$ is a vector of multipliers associated with N^{b} binding constraints and $\hat{y}$ is the vector of bound post-states such that $\partial_{y_i} \bar{\mathrm{g}}_t^{\mathrm{b}}(\bar{m}, y) \neq 0$. As such, we will have:

$$\begin{aligned}(37) \quad & \partial_{y_i} \bar{\mathrm{u}}(\varphi(\tilde{a}), \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}) + \hat{\mu}^\top \partial_{y_i} \bar{\mathrm{g}}_t(\bar{m}, y) \\ & \quad + \mathbb{E}_t \partial_{y_i} \Lambda_{t+1}(x')^\top \partial_{y_i} \mathrm{F}_{t+1}^m \left(d_t, \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}, w_{t+1}\right) = 0, \qquad i \,\mathrm{s.t}\ \partial_{y_i} \bar{\mathrm{g}}_t^{\mathrm{b}}(\bar{m}, y) \neq 0\end{aligned}$$

Next, since $\varphi_0(\tilde{a}) \in U_{0,\bar{m}}$ and $\Psi(\varphi_0(\tilde{a}), \tilde{a}) = 0$ for i such that $\partial_{y_i} \bar{\mathrm{g}}_t^{\mathrm{b}}(\bar{m}, y) = 0$, we will have:

$$\partial_{y_i} \bar{\mathrm{u}}(\varphi(\tilde{a}), \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}) + \mathbb{E}_t \partial_{y_i} \Lambda_{t+1}(x')^\top \partial_{y_i} \mathrm{F}_{t+1}^m \left(d_t, \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}, w_{t+1}\right) = 0 \tag{38}$$

Moreover, since $\left(\varphi(\tilde{a}), \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}\right) \in K_{l,t}$, we also have:

$$\bar{\mathrm{g}}_t^{\mathrm{b}}(\varphi(\tilde{a}), \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}) = 0 \tag{39}$$

Finally, let φ be defined by $\tilde{a} \mapsto \mu, \pi_Y \mathrm{em}_{\tilde{A}} \tilde{a}$, where μ is obtained by permutation of indicies of vectors in $[\hat{\mu}, \mathbf{0}_{N_g - N^{\mathrm{b}}}]$ and $\mathbf{0}_{N_g - N^{\mathrm{b}}}$ is the value of the multipliers for the non-binding constraints. To complete the proof, note that since (37) - (39) hold, we must have $\nabla_a \bar{\mathbf{S}}_t(\varphi_0(\tilde{a}), \varphi(\tilde{a})) = 0$ for each $\tilde{a} \in U_{0,\tilde{a}}$ and such that $(\varphi_0(\tilde{a}), \varphi(\tilde{a})) \in K_{l,t}$. □

Proof of Claim 2. The proof is analogous to the proof of Claim 1. □

4. Proofs for Section 4.2

Note that in the presence of discrete choices, Θ_t may not be defined everywhere on its domain, but the image of Θ_t will have the same dimension as the post-state. In particular, we can verify that the image of the policy functions $\sigma_t(\bar{M}_t)$ will be a union of connected subsets. To do so, let $\bar{M}_t^{\vec{\boldsymbol{d}}_{t+1}}$ denote the set of points in $\bar{M}_t$ that 'lead to' the future sequence of discrete choices $\vec{\boldsymbol{d}}_{t+1}$ and so,

$$\bar{M}^{\mathbf{d}_{t+1}} = \left\{ \bar{m}_t \in \bar{M}_t \,\middle|\, \vec{\boldsymbol{d}}_{t+1} = \underset{\vec{\boldsymbol{d}} \in \mathbf{D}_{t+1}}{\arg\max}\, \bar{\mathrm{v}}_t^{d_t, \vec{\boldsymbol{d}}}(x_t) \right\} \tag{40}$$

where $x_t = (w_t, \bar{m}_t, z_t)$, recall w_t, d_t and z_t are held fixed, and note that $\bar{M}_t = \bigcup_{\vec{\boldsymbol{d}} \in \mathbf{D}_{t+1}} \bar{M}_t^{\mathbf{d}}$.

Proposition 5. *If Assumption E.3 holds, then* $\sigma_t(\bar{M}_t) = \bigcup_{\vec{\boldsymbol{d}} \in \mathbf{D}_{t+1}} \sigma_t(\bar{M}_t^{\vec{\boldsymbol{d}}})$*, where each* $\bar{M}_t^{\vec{\boldsymbol{d}}}$ *is a connected set with equal dimension to* $\sigma_t(\bar{M}_t)$*.*

For the following Lemma, recall that ϵ is the radius of the exogenous grid.

Claim 3. *Let Assumptions E.1 - E.3 hold and let* $m \in \bar{M}_t$ *be such that* $\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m) \in \mathrm{K}_{l,t}^{\mathrm{o}}$ *for some* l*. There exists* $\bar{\bar{\delta}}$ *such that for any* $\bar{\delta} < \bar{\bar{\delta}}$*, there exists* $\bar{N}^{\mathcal{A}}$ *such that if* $|\mathcal{A}_t| > \bar{N}^{\mathcal{A}}$*, then there exists* $m_{t,i}^{\#} \in \mathcal{M}_t$ *with* $m_{t,i}^{\#} \in \mathbb{B}_{\bar{\delta}}(m)$ *and* $m_{t,i}^{\#} \in \Theta_t(a_{t,i}^{\#})$*.*

Proof. By Proposition 5, there exists $\bar{\bar{\delta}}$ such that if $m' \in \mathbb{B}_{\bar{\bar{\delta}}}(m)$, then $\sigma_t(m') = \sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m')$, where $\vec{\boldsymbol{d}}_{t+1}$ is fixed as the optimal future sequence of discrete choices from m. Next, note that $\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}$ is locally a homeomorphism onto its own image (by Theorem 2 or by Assumption I.3). As such, there exists $\bar{\bar{\delta}}$ such that $\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(\mathbb{B}_{\bar{\delta}}(m))$ is contained in an open neighbourhood U in $\mathrm{K}^{\mathrm{o}}_{l,t} \cap \sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(\bar{M}_t)$ for any $\bar{\delta} < \bar{\bar{\delta}}$. By Assumption E.1, for grid size large enough, we must have $a^{\#}_{t,i} \in U \subset \mathrm{K}^{\mathrm{o}}_{t,l} \cap \sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(\bar{M}_t)$. Thus, we must have that $m^{\#}_{t,i} \in \mathbb{B}_{\bar{\delta}}(m)$ and $m^{\#}_{t,i} \in \Theta_t(a^{\#}_{t,i})$ where $m^{\#}_{t,i} \in \mathcal{M}_t$. □

Proposition 6. *Let Assumptions E.1 - E.3 hold, let* $\rho_r > \epsilon L_2$, $\bar{J} = L_1$ *and let* $m \in \bar{M}_t$ *be such that* $\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m) \in \mathrm{K}^{\mathrm{o}}_{l,t}$ *for some* l. *There exists* $\delta_m > 0$ *and* ϵ *small enough such that if* $m^{\#}_{t,i} \in \mathbb{B}_{\delta_m}(m)$, *then* $\|\kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t(m^{\#}_{t,i})) - \kappa^{\vec{\boldsymbol{d}}_{t+1}}(a^{\#}_{t,i})\| < 2\epsilon$.

Proof. By Claim 3, there exists $\delta_m/2$ such that for all ϵ small enough, for some $m^{\#}_{t,k^\star}$, we have $m^{\#}_{t,k^\star} \in \mathbb{B}_{\delta_m/2}(m)$ where $a^{\#}_{t,k^\star}$ is optimal. Moreover, setting $\delta_m/2$ small enough, for any other point $m^{\#}_{t,i}$ with $m^{\#}_{t,i} \in \mathbb{B}_{\delta_m/2}(m)$, $\|m^{\#}_{t,i} - m^{\#}_{t,k^\star}\| < L_2\epsilon$. Let $\boldsymbol{d}_{t+1} = \max_{\vec{\boldsymbol{d}}'} \mathrm{Q}_t^{\vec{\boldsymbol{d}}'}(x^{\#}_{t,k^\star}, d_t, a^{\#}_{t,k^\star})$ where $\vec{\boldsymbol{d}}'$ is feasible and where $x^{\#}_{t,k^\star} = z_t, m^{\#}_{t,k^\star}$.

By Proposition 5, taking $\delta_m/2$ small enough, we have that $\boldsymbol{d}_{t+1}$ is the optimal future sequence of discrete choices for any other $m^{\#}_{t,i} \in \mathbb{B}_{\delta_m/2}(m)$. However, let $\hat{\vec{\boldsymbol{d}}}_{t+1}$ be such that $a^{\#}_{t,i} = \sigma_t^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(m^{\#}_{t,i})$ and suppose that $\hat{\vec{\boldsymbol{d}}}_{t+1} \neq \boldsymbol{d}_{t+1}$. Since $L_2\epsilon < \rho_r$, we have $\|m^{\#}_{t,i} - m^{\#}_{t,k^\star}\| \leq \rho_r$. Next, letting $\underline{\kappa} = \|\kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t(m^{\#}_{t,i})) - \kappa^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(a^{\#}_{t,i})\|$, we have

$$\begin{aligned}
\|\kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m^{\#}_{t,k^\star})) - \kappa^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(\sigma_t^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(m^{\#}_{t,i}))\| &\geq \Big|\|\kappa^{\boldsymbol{d}_{t+1}}(\sigma_t^{\boldsymbol{d}_{t+1}}(m^{\#}_{t,i})) - \kappa^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(\sigma_t^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(m^{\#}_{t,i}))\| \\
&\quad - \|\kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m^{\#}_{t,i})) - \kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m^{\#}_{t,k^\star}))\|\Big| \\
&\geq |\underline{\kappa} - \epsilon|
\end{aligned}$$

Finally, letting $\underline{\kappa} > \epsilon$, we have

$$\frac{\underline{\kappa}}{\|m^{\#}_{t,i} - m^{\#}_{t,k^\star}\|} < \frac{\|\kappa^{\vec{\boldsymbol{d}}_{t+1}}(\sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(m^{\#}_{t,k^\star})) - \kappa^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(\sigma_t^{\hat{\vec{\boldsymbol{d}}}_{t+1}}(m^{\#}_{t,i}))\| + \epsilon}{\|m^{\#}_{t,i} - m^{\#}_{t,k^\star}\|} < \bar{J} + \frac{\epsilon}{\|m^{\#}_{t,i} - m^{\#}_{t,k^\star}\|} \tag{41}$$

which will imply $\underline{\kappa} < \bar{J}L^{-1}\epsilon + \epsilon$. Letting $\bar{J} = L$ allows us to arrive at the result. □

Proof of Theorem 3. Let $\bar{m}_t \in \bar{M}_t$ satisfy the conditions of the Theorem. By Proposition 5, there exists δ_1 such that if $m' \in \mathbb{B}_{\delta_1}(\bar{m}_t)$, then $\sigma_t(m') = \sigma_t^{\vec{\boldsymbol{d}}_{t+1}}(\bar{m}_t)$, where $\vec{\boldsymbol{d}}_{t+1}$ is fixed as the optimal future sequence of discrete choices from $\bar{m}_t$.

Next, let $\mathbb{S}^\star \subset \mathbb{B}_{\delta_1}(\bar{m}_t)$ be a $N_{\bar{M}}$ simplex (not neccesarily contained in $\mathcal{M}_t$). Let $\nu_0, \ldots, \nu_{N_{\bar{M}}+1}$ be the set of vertices of $\mathbb{S}^\star$. There exists $\bar{\delta}$ such if $\bar{\nu}_i \in \mathbb{B}_{\bar{\delta}}(\nu_i)$, then $\{\bar{\nu}_0, \ldots, \bar{\nu}_{N_{\bar{M}}+1}\}$ is a $N_{\bar{M}}$ simplex. However, by Claim 3, there exists $|\mathcal{A}_t|$ large enouch such that for each $\bar{\nu}_i$, there exists $\bar{m}^{\#}_{t,i} \in \mathcal{M}_t$ such that $\bar{m}^{\#}_{t,i} \in \Theta_t(a^{\#}_{t,i})$ and $\bar{m}^{\#}_{t,i} \in \mathbb{B}_{\bar{\delta}}(\nu_i)$. Let $\bar{\mathbb{S}}$ be the simplex defined by these $\left\{\bar{m}^{\#}_{t,0}, \ldots, \bar{m}^{\#}_{t,i}, \ldots \bar{m}^{\#}_{t,N_{\bar{M}}+1}\right\}$. We must have $\mathbb{S} \subset \mathbb{B}_{\delta_1}(m)$.

Next, by Proposition 6, for a δ_1 small enough, there exists ϵ arbitrarily small such that if $\bar{m}^{\#}_{t,i} \in \mathbb{B}_{\delta_{\bar{m}_t}}(\bar{m}_t)$, then $\|\kappa(\sigma_t(\bar{m}^{\#}_{t,i})) - \kappa(a^{\#}_{t,i})\| < 2\epsilon$. Moreover, $\bar{m}^{\#}_{t,i}$ is outside the search radius of any other optimal point not leading to the same future sequence of discrete choices. Thus, $\mathbb{S}^\star \subset \mathcal{M}_t^{\mathrm{RFC}}$. Apply Theorem Stämpfle (2000), the maximum approximation error between will be ϵ. Since the error between the interpolant and the true value function is 2ϵ, the result follows. □

5. Additional Supporting Analysis

For completeness, this section collects additional, standard, results required for the proofs in this paper. For the following proposition, consider an underlying probabilty space $(\Omega, \mathscr{F}, \mathbb{P})$.

Proposition 7. *Let $\boldsymbol{X} = L^2(\Omega, \mathbb{R}^N)$ and $\boldsymbol{\phi} \in \boldsymbol{X}$. For a fixed $\boldsymbol{x} \in \boldsymbol{X}$, suppose that for all $\boldsymbol{\nu} \in \boldsymbol{X}$,*

$$\langle \boldsymbol{\phi}, \boldsymbol{\nu} - \boldsymbol{x} \rangle \geq \int \mathrm{f}(\boldsymbol{\nu}) - \mathrm{f}(\boldsymbol{x})\, d\,\mathbb{P},$$

where $\mathrm{f} : \mathbb{R}^N \to \mathbb{R}_{-\infty}$ and $\mathrm{f}(\boldsymbol{x}) \in L^2(\Omega, \mathbb{R})$. Then, for any $\boldsymbol{\nu} \in \boldsymbol{X}$,

$$\boldsymbol{\phi}(\boldsymbol{\nu} - \boldsymbol{x}) \geq \mathrm{f}(\boldsymbol{\nu}) - \mathrm{f}(\boldsymbol{x}) \quad a.e.$$

Proof. Assume, for the sake of contradiction, that there exists and $\tilde{\boldsymbol{\nu}} \in \boldsymbol{X}$ and a non-null set $A \subseteq \Omega$ with strictly positive measure such that for all $\omega \in A$,

$$\boldsymbol{\phi}(\tilde{\boldsymbol{\nu}}(\omega) - \boldsymbol{x}(\omega)) < \mathrm{f}(\tilde{\boldsymbol{\nu}}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega))$$

Define a function $\boldsymbol{\nu} : \Omega \to \mathbb{R}^N$ by

$$\boldsymbol{\nu}(\omega) = \begin{cases} \tilde{\boldsymbol{\nu}}(\omega) & \text{if } \omega \in A, \\ \boldsymbol{x}(\omega) & \text{otherwise,} \end{cases}$$

where $\tilde{\boldsymbol{\nu}}(\omega)$ is chosen such that the strict inequality above is satisfied for $\omega \in A$.

By the construction of $\boldsymbol{\nu}$, $\boldsymbol{\nu}(\omega) - \boldsymbol{x}(\omega) = 0$ for $\omega \notin A$ and $\boldsymbol{\nu}(\omega) - \boldsymbol{x}(\omega) = \tilde{\boldsymbol{\nu}}(\omega) - \boldsymbol{x}(\omega)$ for $\omega \in A$. Therefore,

$$\langle \boldsymbol{\phi}, \boldsymbol{\nu} - \boldsymbol{x} \rangle = \int_A \boldsymbol{\phi}(\tilde{\boldsymbol{\nu}}(\omega) - \boldsymbol{x}(\omega))\, \mathbb{P}(d\omega),$$

Given the strict inequality on A, integrating over A yields

$$\langle \boldsymbol{\phi}, \boldsymbol{\nu} - \boldsymbol{x} \rangle < \int_A (\mathrm{f}(\tilde{\boldsymbol{\nu}}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega)))\, \mathbb{P}(d\omega), = \int \mathrm{f}(\boldsymbol{\nu}) - \mathrm{f}(\boldsymbol{x})\, d\mathbb{P},$$

which contradicts the assumption that

$$\langle \boldsymbol{\phi}, \boldsymbol{\nu} - \boldsymbol{x} \rangle \geq \int (\mathrm{f}(\boldsymbol{\nu}) - f(\boldsymbol{x}))\, dP \quad \text{for all } \boldsymbol{\nu} \in \boldsymbol{X}.$$

Thus, our contradictory assumption must be false, implying that there cannot exist a non-null set A where the inequality is violated. Consequently, it holds that

$$\boldsymbol{\phi}(\boldsymbol{\nu}(\omega) - \boldsymbol{x}(\omega)) \geq \mathrm{f}(\boldsymbol{\nu}(\omega)) - \mathrm{f}(\boldsymbol{x}(\omega))$$

almost everywhere, as was to be shown. □

References

Hernandez-Lerma, O. and Lasserre, J. (2012). *Discrete-time Markov control processes: Basic optimality criteria.* Stochastic modelling and applied probability. Springer New York.

Kumagai, S. (1980). An implicit function theorem: Comment. *Journal of Optimization Theory and Applications*, 31(2):285–288.

Luenberger, D. (1997). *Optimization by vector space methods*. Professional Series. Wiley.

Shanker, A. (2017). Existence of recursive constrained optima in the heterogenous agent neoclassical growth model. *Working Paper*.

Sorger, G. (2015). *Dynamic economic analysis: Deterministic models in discrete time*. Cambridge University Press.

Stämpfle, M. (2000). Optimal estimates for the linear interpolation error on simplices. *Journal of Approximation Theory*, 103(1):78–90.