Submitted to *Econometrica*

SUPPLEMENT TO "INSURANCE AND INEQUALITY WITH PERSISTENT PRIVATE INFORMATION"

ALEXANDER W. BLOEDEL
Department of Economics, UCLA

R. VIJAY KRISHNA
Department of Economics, Florida State University

OKSANA LEUKHINA
Research Division, Federal Reserve Bank of St. Louis

In this Supplemental Appendix (henceforth SA), we prove Theorems 1 and 2 and associated results, and discuss Condition R.5. SA-C proves Theorem 1 and Proposition 4.4. SA-D proves Corollary 4.1. SA-E presents the proof of Theorem 2, an important step of which is proving Theorem 3(d). SA-F and SA-G collect facts about the first-best and pathwise properties of Markov chains, respectively. SA-H discusses Condition R.5. While this SA is mostly self-contained, some auxiliary results are proved in Bloedel, Krishna, and Leukhina (2024).

APPENDIX C: PROOF OF THEOREM 1

Herein, we assume that the environment is (TVC)-Regular (as in the statement of Theorem 1). SA-C.1 presents the Lagrangian and first-order optimality conditions for the Bellman equation (FE). SA-C.2 proves Proposition 4.4 (Step 1 in the sketch from Section 4.3). SA-C.3 presents intermediate steps towards the proof of Theorem 1 (most of Step 2 in the sketch). SA-C.4 presents the main convergence proofs (most of Step 3 in the sketch).

C.1. *Optimality Conditions*

Recall that the set of *recursive constraints* consists of the *promise keeping* constraints

$$v_i = u_i + \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i] \qquad (\mathrm{PK}_i)$$

for all $i \in S$, and the *incentive compatibility* constraints

$$u_i + \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i] \geq \psi(u_j, i, j) + \alpha\, \mathsf{E}^{\mathbf{f}_i}\left[\mathbf{w}_j\right] \qquad (\mathrm{IC}_{ij})$$

for all $i, j \in S$ with $i > j$ (per Assumption NHB).[1]

Under (TVC)-Regularity, Proposition 3.2 reduces the principal's problem to a family of smooth, strictly convex, finite-dimensional minimization problems. Thus, under Condition R.3, standard results imply that optimal menus in (FE) can be characterized via saddle points of a Lagrangian function (see, e.g., Exercise 7 on p. 236 and Theorem 2 on p. 221 of Luenberger (1969)). Letting $\lambda_i \in \mathbb{R}$ denote a multiplier on the promise keeping constraint (PK$_i$) and $\mu_{ij} \geq 0$

Alexander W. Bloedel: abloedel@econ.ucla.edu
R. Vijay Krishna: rvk3570@gmail.com
Oksana Leukhina: oksana.m.leukhina@gmail.com

[1]The incentive constraints are written here in a slightly different, but equivalent, form than in Section 3.

denote a multiplier on the incentive constraint (IC_{ij}), the Lagrangian for this problem is

$$\mathscr{L}(\mathbf{v}, s, \mathbf{u}, \mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^{d} f_{si}\big(C(u_i, i) + \alpha P(\mathbf{w}_i, i)\big) + \sum_{i=1}^{d} \lambda_i\big(v_i - u_i - \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i]\big)$$
$$- \sum_{i=2}^{d}\sum_{j=1}^{i-1} \mu_{ij}\big(u_i + \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i] - \psi(u_j, i, j) - \alpha\, \mathsf{E}^{\mathbf{f}_i}\big[\mathbf{w}_j\big]\big).$$

For notational ease, we henceforth extend μ_{ij} to all pairs $i, j \in \mathbb{N}$, with the convention that $\mu_{ij} = 0$ if $j \geq i$, $i \notin S$, or $j \notin S$.

The necessary and sufficient optimality equations consist of the envelope conditions

$$P_i(\mathbf{v}, s) = \lambda_i(\mathbf{v}, s) \tag{Env$_i$}$$

for all $i \in S$, the first-order conditions for flow utilities

$$f_{si} C'(u_i(\mathbf{v}, s), i) = \lambda_i(\mathbf{v}, s) + \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{v}, s) - \sum_{k=i+1}^{d} \psi'(u_i(\mathbf{v}, s), k, i)\mu_{ki}(\mathbf{v}, s) \tag{FOCu_i}$$

for all $i \in S$, and the first-order conditions for contingent continuation utilities

$$f_{si} P_j(\mathbf{w}_i(\mathbf{v}, s), i) = f_{ij}\left(\lambda_i(\mathbf{v}, s) + \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{v}, s)\right) - \sum_{k=i+1}^{d} f_{kj}\mu_{ki}(\mathbf{v}, s) \tag{FOC$\mathbf{w}_{ij}$}$$

for all $i, j \in S$ with $i > j$ (per Assumption NHB), where $(u_i(\mathbf{v}, s), \mathbf{w}_i(\mathbf{v}, s))_{i \in S} \in \Gamma(\mathbf{v})$ is the optimal menu at state $(\mathbf{v}, s)$ and $\lambda_i(\mathbf{v}, s)$ and $\mu_{ij}(\mathbf{v}, s)$ are the corresponding multipliers.[2]

By Proposition 3.2, the policy functions characterized by the above optimality conditions induce the (unique) optimal contract. Accordingly, we henceforth let ξ denote the optimal contract, let $\xi^f(\mathbf{v}, s, i) := u_i(\mathbf{v}, s)$ denote the *flow* utility policy functions, and let $\xi^c(\mathbf{v}, s, i) := \mathbf{w}_i(\mathbf{v}, s)$ denote the *continuation* utility policy functions. Going forward, we will utilize both of these notational conventions for the policy functions, depending on which one is more convenient for the task at hand.

C.2. *Proof of Proposition 4.4*

By Proposition 3.2(b), the value function $P(\cdot, s) \in \mathbf{C}^1(D)$. Hence, the directional derivative $\mathsf{D}_\mathbf{1} P(\cdot, s) = \sum_{i \in S} P_i(\cdot, s)$ and is real-valued on D. For each $t \in \mathbb{N}$, integrability of the random variable $\mathsf{D}_\mathbf{1} P(\mathbf{v}^{(t)}, s^{(t)})$ then follows from finiteness of S and positivity of the directional derivative (established in Lemma C.1 below). For the martingale property, let $(\mathbf{v}, s) \in D \times S$ be given. Summing the (Env$_i$) over $i \in S$ delivers

$$\mathsf{D}_\mathbf{1} P(\mathbf{v}, s) = \sum_{i=1}^{d} \lambda_i(\mathbf{v}, s). \tag{C.1}$$

[2]We omit the usual complementary slackness conditions.

For each fixed $i \in S$, summing the (FOC$\mathbf{w}_{ij}$) over $j \in S$ yields

$$f_{si} \cdot \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_i(\mathbf{v},s),i) = \lambda_i(\mathbf{v},s) + \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{v},s) - \sum_{k=i+1}^{d} \mu_{ki}(\mathbf{v},s). \tag{C.2}$$

Now, summing (C.2) over $i \in S$ and noting that $\sum_{i=1}^{d} \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{v},s) = \sum_{i>k\in S} \mu_{ik}(\mathbf{v},s) = \sum_{i=1}^{d} \sum_{k=i+1}^{d} \mu_{ki}(\mathbf{v},s)$ delivers $\sum_{i\in S} f_{si} \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_i(\mathbf{v},s),i) = \sum_{i=1}^{d} \lambda_i(\mathbf{v},s)$. Combined with (C.1), this delivers the martingale property $\mathsf{D}_{\mathbf{1}} P(\mathbf{v},s) = \sum_{i=1}^{d} f_{si} \mathsf{D}_{\mathbf{1}} P(\mathbf{w}_i(\mathbf{v},s),i)$.

The following lemma then completes the proof:

LEMMA C.1: *For every $s \in S$, the directional derivative $\mathsf{D}_{\mathbf{1}} P(\cdot,s)$ is strictly positive.*

PROOF: We first show that each $\mathsf{D}_{\mathbf{1}} P(\cdot,s)$ is non-negative on D, i.e., that $P(\cdot,s)$ is non-decreasing in direction $\mathbf{1}$. Lemma F.1 shows that the first-best value function Q^* is non-decreasing in this direction. We will show that P inherits this property from Q^*. The proof is order-theoretic. Let $[Q^*, P]$ denote the order interval (in the pointwise order) of functions $Q : D \times S \to \mathbb{R}$ that lie weakly above Q^* and weakly below P. Let $\Phi := \{Q \in [Q^*, P] : Q(\mathbf{v} + \varepsilon\mathbf{1}, s) \geqslant Q(\mathbf{v},s)\ \forall\, \mathbf{v} \in D,\ \varepsilon > 0 \text{ s.t. } \mathbf{v} + \varepsilon\mathbf{1} \in D\}$ denote the subset of functions in $[Q^*, P]$ that are non-decreasing in direction $\mathbf{1}$. By Lemma F.1, $Q^* \in \Phi$.

CLAIM C.2: *Φ is a lattice in the pointwise order.*

PROOF OF CLAIM C.2: Let $f, g \in \Phi$ be given. Clearly, $f \vee g, f \wedge g \in [Q^*, P]$. Now, fix $(\mathbf{v},s) \in D \times S$ and $\varepsilon > 0$ such that $\mathbf{v}' := \mathbf{v} + \varepsilon\mathbf{1} \in D$. If f and g are ordered the same way at $(\mathbf{v},s)$ and $(\mathbf{v}',s)$, we are done. So suppose, without loss of generality, that $f(\mathbf{v},s) \geq g(\mathbf{v},s)$ and $g(\mathbf{v}',s) \geq f(\mathbf{v}',s)$. Then, $(f \wedge g)(\mathbf{v}',s) = f(\mathbf{v}',s) \geqslant f(\mathbf{v},s) \geq (f \wedge g)(\mathbf{v},s)$. Similarly, $(f \vee g)(\mathbf{v}',s) \geq f(\mathbf{v}',s) \geqslant f(\mathbf{v},s) = (f \vee g)(\mathbf{v},s)$. Thus, $f \vee g$, $f \wedge g \in \Phi$ as desired. *Q.E.D.*

CLAIM C.3: *The lattice Φ is complete.*

PROOF OF CLAIM C.3: Let $F \subseteq \Phi$ be nonempty and define $\overline{f}(\mathbf{v},s) := \sup_{f\in F} f(\mathbf{v},s)$ and $\underline{f}(\mathbf{v},s) := \inf_{f\in F} f(\mathbf{v},s)$ for each $(\mathbf{v},s) \in D \times S$. Clearly, we have $\overline{f}, \underline{f} \in [Q^*, P]$. We show that $\overline{f}$ is non-decreasing in direction $\mathbf{1}$, and hence $\overline{f} \in \Phi$ (the proof that $\underline{f} \in \Phi$ is symmetric). Suppose towards a contradiction that there exists $(\mathbf{v},s) \in D \times S$ and some $\varepsilon > 0$ such that $(\mathbf{v}',s) \in D \times S$, where $\mathbf{v}' = \mathbf{v} + \varepsilon\mathbf{1}$, and $\overline{f}(\mathbf{v},s) > \overline{f}(\mathbf{v}',s)$. Then $\overline{f}(\mathbf{v},s) - \delta \geq \overline{f}(\mathbf{v}',s)$ for some $\delta > 0$. By definition of $\overline{f}$, there exists an $f \in F$ such that $f(\mathbf{v},s) > \overline{f}(\mathbf{v},s) - \delta$. Combining these inequalities and the definition of $\overline{f}$ yields $f(\mathbf{v},s) > \overline{f}(\mathbf{v}',s) \geq f(\mathbf{v}',s)$, which contradicts that $f \in F \subseteq \Phi$. We conclude that $\overline{f} \in \Phi$, as desired. *Q.E.D.*

Let $\overline{\mathbb{R}}$ denote the extended reals, and let $\overline{\mathbb{R}}^{D\times S}$ denote the space of functions $f : D \times S \to \overline{\mathbb{R}}$. Define the Bellman operator $T : \overline{\mathbb{R}}^{D\times S} \to \overline{\mathbb{R}}^{D\times S}$ by

$$TQ(\mathbf{v},s) := \inf_{(u_i,\mathbf{w}_i)_{i\in S} \in \Gamma(\mathbf{v})} \sum_{i=1}^{d} f_{si}\left[C(u_i,i) + \alpha Q(\mathbf{w}_i,i)\right]. \tag{T}$$

CLAIM C.4: *$T : \Phi \to \Phi$ is well-defined and monotone.*[3]

[3]That is, the image $T(\Phi) \subseteq \Phi$ and, if $Q, \hat{Q} \in \Phi$ satisfy $Q \leq \hat{Q}$, then $TQ \leq T\hat{Q}$.

PROOF OF CLAIM C.4: It is easy to see that the map $T : \Phi \to \overline{\mathbb{R}}^{D\times S}$ is monotone. It remains to show that the image $T(\Phi) \subseteq \Phi$. Let $Q \in \Phi$ be given. Since $Q \leq P$ by definition and $P = TP$ by Proposition 3.2, the monotonicity of T implies that $TQ \leq P$. Since $Q^* \leq Q$ by definition and $Q^* \leq TQ^*$ by Lemma F.1,[4] the monotonicity of T implies that $Q^* \leq TQ$. Thus, $TQ \in [Q^*, P]$. To show that $TQ \in \Phi$, let $(\mathbf{v}, s) \in D \times S$ and $\varepsilon > 0$ such that $\mathbf{v} - \varepsilon\mathbf{1} \in D$ be given. For every $\delta > 0$, there exists $(u_i^\delta, \mathbf{w}_i^\delta)_{i\in S} \in \Gamma(\mathbf{v})$ attaining within δ of the infimal payoff in (T). Then $\delta + TQ(\mathbf{v}, s) \geq \sum_{i\in S} f_{si}\big[C(u_i^\delta, i) + \alpha Q(\mathbf{w}_i^\delta, i)\big] \geq \sum_{i\in S} f_{si}\big[C(u_i^\delta, i) + \alpha Q(\mathbf{w}_i^\delta - \frac{\varepsilon}{\alpha}\mathbf{1}, i)\big] \geq TQ(\mathbf{v} - \varepsilon\mathbf{1}, s)$ where the first inequality is by δ-optimality of the menu, the second inequality is by $Q \in \Phi$, and the third inequality is because $(u_i^\delta, \mathbf{w}_i^\delta - \frac{\varepsilon}{\alpha}\mathbf{1})_{i\in S} \in \Gamma(\mathbf{v} - \varepsilon\mathbf{1})$. Sending $\delta \to 0$ yields $TQ(\mathbf{v}, s) \geq TQ(\mathbf{v} - \varepsilon\mathbf{1}, s)$. We conclude that $TQ \in \Phi$. Thus, the map $T : \Phi \to \Phi$ is well-defined. *Q.E.D.*

Given Claims C.2, C.3, and C.4, Tarski's Fixed Point Theorem implies that T has a smallest fixed point in Φ. This smallest fixed point must be P, since P is the smallest fixed point of $T : \overline{\mathbb{R}}^{D\times S} \to \overline{\mathbb{R}}^{D\times S}$ that lies pointwise above Q^* (Proposition 3.2) and $Q^* \in \Phi$. Hence, $P \in \Phi$. We conclude that $\mathsf{D}_\mathbf{1} P(\cdot, s) \geq 0$ on D for each $s \in S$, as desired.

To complete the proof, we now show that the directional derivative is strictly positive. Suppose, towards a contradiction, that there exists $(\mathbf{v}, s) \in D \times S$ such that $\mathsf{D}_\mathbf{1} P(\mathbf{v}, s) = 0$. For this fixed $(\mathbf{v}, s)$, define the function $g : [-\bar{\varepsilon}, \bar{\varepsilon}] \to \mathbb{R}$ by $g(y) := P(\mathbf{v} + y\mathbf{1}, s)$, where $\bar{\varepsilon} > 0$ is chosen sufficiently small that $\mathbf{v} + y\mathbf{1} \in D$ for all for $y \in [-\bar{\varepsilon}, \bar{\varepsilon}]$. (Such $\bar{\varepsilon} > 0$ exist by Theorem 3(a).) By construction, $g'(y) = \mathsf{D}_\mathbf{1} P(\mathbf{v} + y\mathbf{1}, s) \geq 0$ for all such y. At the same time, $g'(\cdot)$ is strictly increasing because P is strictly convex (Proposition 3.2(b)). Therefore, $g'(0) = \mathsf{D}_\mathbf{1} P(\mathbf{v}, s) = 0$ requires that $g'(y) < 0$ for $y < 0$, delivering the desired contradiction. We conclude that $\mathsf{D}_\mathbf{1} P(\cdot, s) > 0$ on D, as desired. *Q.E.D.*

C.3. *Intermediate Steps Towards the Proof of Theorem 1*

This SA consists of several parts and culminates in Lemma C.18, which shows that the marginal cost martingale necessarily "splits" after consecutive realizations of the highest type d. To this end, SA-C.3.1 presents preliminary facts about the efficiency problem (Eff_i) from Section 4.3. SA-C.3.2 shows that the optimal contract is efficient (i.e., solves (Eff_i)) after consecutive d-type realizations, and records important properties of the marginal cost martingale at such histories. SA-C.3.3 studies an "interim" reformulation of the Bellman equation (FE), which lets us relate policy functions and optimal Lagrange multipliers across different values of the previous report $s \in S$. Finally, SA-C.3.4 uses the preceding results and facts about the first-best solution (recorded in SA-F) to prove Lemma C.18.

In what follows, we will make repeated use of the following fact:

LEMMA C.5: *Let $Y_s := \mathsf{D}P(D, s) \subseteq \mathbb{R}^d$ denote the image of D under the derivative map $\mathsf{D}P(\cdot, s) : D \to \mathbb{R}^d$. For every $s \in S$, the map $\mathsf{D}P(\cdot, s) : D \to Y_s$ is a homeomorphism.*

PROOF: Since $P(\cdot, s)$ is strictly convex, $\mathsf{D}P(\cdot, s) : D \to \mathbb{R}^d$ is injective.[5] Then, since $D \subseteq \mathbb{R}^d$ is open (Theorem 3(a)) and $\mathsf{D}P(\cdot, s)$ is continuous (Proposition 3.2(b)), Brouwer's Invariance of Domain Theorem (e.g., Hatcher, 2001, Theorem 2B.3) implies that $\mathsf{D}P(\cdot, s) : D \to \mathbb{R}^d$ is an open map. Hence, the bijection $\mathsf{D}P(\cdot, s) : D \to Y_s$ is a homeomorphism. *Q.E.D.*

[4]In particular, Lemma F.1 shows that Q^* satisfies the Bellman equation (F.1), in which the feasible set $\Gamma(\mathbf{v})$ from (T) is replaced by the larger feasible set $\Gamma^{\text{FB}}(\mathbf{v})$ (which omits incentive constraints). This implies that $Q^* \leq TQ^*$, because infimizing over a smaller feasible set can only increase the principal's costs.

[5]Strict convexity implies that the derivative is *strictly monotone*: for all distinct $\mathbf{v}, \mathbf{v}' \in D$, $\langle \mathbf{v}' - \mathbf{v}, \mathsf{D}P(\mathbf{v}', s) - \mathsf{D}P(\mathbf{v}, s)\rangle > \mathbf{0}$. Clearly, if $\mathsf{D}P(\cdot, s)$ were not injective, strict monotonicity would fail.

C.3.1. *The Efficiency Problem*

Recall the *efficiency problem* (Eff$_i$) from Section 4.3, re-stated here for convenience:

$$K(w,i) := \min_{\mathbf{w}_i \in D} P(\mathbf{w}_i, i) \qquad \text{s.t.} \qquad \mathbf{E}^{\mathbf{f}_i}[\mathbf{w}_i] \geq w. \tag{Eff$_i$}$$

LEMMA C.6: *For every $i \in S$, the following properties hold:*

(a) *For each $w \in \mathcal{U}$, (Eff$_i$) has a unique solution $\mathbf{w}^\dagger(w,i)$ and $w = \mathbf{E}^{\mathbf{f}_i}[\mathbf{w}^\dagger(w,i)]$.*

(b) *The policy function $\mathbf{w}^\dagger(\cdot,i) : \mathcal{U} \to D$ is continuous and injective, and thus defines a bijection between $\mathcal{U}$ and its image $E_i := \{\mathbf{v} \in D : \exists w \in \mathcal{U} \text{ s.t. } \mathbf{v} = \mathbf{w}^\dagger(w,i)\}$.*

(c) *The value function $K(\cdot,i) : \mathcal{U} \to \mathbb{R}$ is well-defined, strictly increasing, strictly convex, continuously differentiable, unbounded above, and satisfies the Inada conditions $\lim_{w\to-\infty} K'(w,i) = 0$ and $\lim_{w\to 0} K'(w,i) = +\infty$.*

PROOF: For part (a), existence follows from routine arguments analogous to those used to establish existence of optimal contracts in Proposition 3.2 (cf. Lemma J.9 in Section J of Bloedel, Krishna, and Leukhina (2024)), uniqueness follows from the strict convexity of $P(\cdot,i)$, and the binding constraint follows from Lemma C.1. For part (b), continuity follows from uniqueness and an application of Berge's Theorem (cf. Lemma J.9 in Section J of Bloedel, Krishna, and Leukhina (2024)). Since the constraint in (Eff$_i$) binds, $\mathbf{w}^\dagger(\cdot,i)$ is clearly injective. Thus, $\mathbf{w}^\dagger(\cdot,i) : \mathcal{U} \to E_i$ is a bijection. For part (c), $K(\cdot,i)$ is finite-valued (hence, well-defined) and strictly convex because P satisfies these properties (Proposition 3.2), strictly increasing by Lemma C.1, and inherits continuously differentiability from $P(\cdot,i)$ via a standard envelope argument for smooth convex problems.[6] It remains to show that $K(\cdot,i)$ is unbounded above and satisfies the claimed Inada conditions. These follow from analogous properties of the value function $K^*(w,i)$ for the first-best analogue of (Eff$_i$), in which P is replaced by Q^* (see (Eff$_i^{\text{FB}}$) in SA-F for the formal definition). Since $Q^* \leq P$ on $D \times S$, clearly $K^* \leq K$ on $\mathcal{U} \times S$. Lemma F.2 shows that $\lim_{w\to 0} K^*(w,i) = +\infty$. If $K(\cdot,i)$ were bounded above, i.e., $\lim_{w\to 0} K(w,i) < +\infty$, then there would exist some $v \in \mathcal{U}$ such that $K^*(v,i) > K(v,i)$, a contradiction. Thus, $\lim_{w\to 0} K(w,i) = +\infty$ and therefore $\lim_{w\to 0} K'(w,i) = +\infty$. Next, Lemma F.2 shows that $\lim_{w\to-\infty} (K^*)'(w,i) = 0$. If $K(\cdot,i)$ were to satisfy $\lim_{w\to-\infty} K'(w,i) > 0$, then there would exist some $v \in \mathcal{U}$ such that $K^*(v,i) > K(v,i)$, again a contradiction. Thus, $\lim_{w\to-\infty} K'(w,i) = 0$. *Q.E.D.*

The efficiency problem (Eff$_i$) admits a Lagrangian $\mathcal{L}^E(w,i,\mathbf{w},\zeta) = P(\mathbf{w},i) - \zeta \cdot \big(\mathbf{E}^{\mathbf{f}_i}[\mathbf{w}] - w\big)$ where $\zeta \geq 0$. By Lemma C.6, the unique solution to (Eff$_i$) at $w \in \mathcal{U}$ is characterized by the first-order and envelope conditions

$$P_j(\mathbf{w}^\dagger(w,i),i) = \zeta(w,i) f_{ij} \tag{FOC$_j$-Eff$_i$}$$

$$K'(w,i) = \zeta(w,i) \tag{Env$_j$-Eff$_i$}$$

where $\zeta(w,i) > 0$ denotes the optimal multiplier. Since Lemma C.6(c) implies that the image $K'(\mathcal{U},i) = \mathbb{R}_{++}$, these optimality conditions imply that the image of the *efficient set* $E_i =$

[6] For instance, part (a) and $P(\cdot,i) \in \mathbf{C}^1(D)$ imply, via the necessary and sufficient first-order condition stated below as (FOC$_j$-Eff$_i$), that there is a unique Lagrange multiplier $\zeta(w,i) \in \mathbb{R}_+$ for (Eff$_i$) at every $w \in \mathcal{U}$. A simple adaptation of Milgrom and Segal (2002, Corollary 5) then delivers $K(\cdot,i) \in \mathbf{C}^1(\mathcal{U})$.

$\{\mathbf{v} \in D : \exists w \in \mathcal{U} \text{ s.t. } \mathbf{v} = \mathbf{w}^\dagger(w,i)\}$ (as defined in Lemma C.6(b)) under the derivative map $\mathsf{D}P(\cdot,i)$ is given by the *efficiency ray*[7]

$$\mathsf{D}P(E_i,i) = \tilde{E}_i := \left\{(P_1,\dots,P_d) \in \mathbb{R}^d_{++} : \frac{P_1}{f_{i1}} = \cdots = \frac{P_d}{f_{id}}\right\}. \tag{$\tilde{\mathrm{E}}_i$}$$

Moreover, by summing the first-order conditions (FOC_j-Eff_i) over $j \in S$ and combining with the envelope condition (Env_j-Eff_i), we obtain

$$K'(w,i) = \mathsf{D}_1 P(\mathbf{w}^\dagger(w,i),i). \tag{C.3}$$

REMARK C.7—Efficiency in the i.i.d. Case: When types are i.i.d., the optimal contract satisfies $\xi^c(\cdot,\cdot,i) \in E_i$ for all $i \in S$. This follows mechanically from the absence of Markov information rents (i.e., $\mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i] = \mathsf{E}^{\mathbf{f}_j}[\mathbf{w}_i]$ for all $i,j \in S$), the Bellman equation (FE) in Proposition 3.2, and the definition of (Eff_i). (It can also be seen by comparing ($\tilde{\mathrm{E}}_i$) to the optimality conditions in SA-C.1.) Moreover, $P(\cdot,i) = P(\cdot,j)$, $K(\cdot,i) = K(\cdot,j)$, and $E_i = E_j$ for all $i,j \in S$, so we can dispense with these type indices. It follows that the $\mathbf{v}^{(t)}$ process evolves in a subset $E \subset D$ that is in bijection to $\mathcal{U} \subset \mathbb{R}$ (Lemma C.6(b)). In effect, E is the image in D of the one-dimensional state space $\mathcal{U}$ from Thomas and Worrall (1990), $K(\cdot)$ is their value function, and $K'(\cdot)$ is their martingale (under NHB).

C.3.2. *Special Properties after d-Type Reports*

LEMMA C.8: *The optimal contract is efficient after type d reports. That is, for every $(\mathbf{v},s) \in D \times S$, we have $\xi^c(\mathbf{v},s,d) \in E_d$.*

PROOF: This follows directly from the Bellman equation (FE) in Proposition 3.2, the fact that the variable $\mathbf{w}_d \in D$ does not enter into any of the (IC_{ij}) constraints for $i \neq d$, and the definition of (Eff_i) (for $i = d$). (Alternatively, the optimality conditions ($\mathrm{FOC}\mathbf{w}_{ij}$) for $i = d$ imply that $\mathsf{D}P(\xi^c(\mathbf{v},s,d),d) \in \tilde{E}_d$, so that ($\tilde{\mathrm{E}}_i$) and Lemma C.5 deliver $\xi^c(\mathbf{v},s,d) \in E_d$.) *Q.E.D.*

LEMMA C.9: *Given any $(\mathbf{v},s) \in D \times S$ and $i \in S$, define $\mathbf{w}_d := \xi^c(\mathbf{v},s,d)$ and $\tilde{\mathbf{w}}_i := \xi^c(\mathbf{w}_d,d,i)$. The following property holds:*

$$\mathsf{D}_1 P(\tilde{\mathbf{w}}_i,i) = \mathsf{D}_1 P(\mathbf{w}_d,d) + \frac{1}{f_{di}}\sum_{k=1}^{i-1}\mu_{ik}(\mathbf{w}_d,d) - \frac{1}{f_{di}}\sum_{k=i+1}^{d}\mu_{ki}(\mathbf{w}_d,d). \tag{MS_i}$$

PROOF: Let $(\mathbf{v},s) \in D \times S$ be given. We begin with the $i = d$ case. The optimality condition ($\mathrm{FOC}\mathbf{w}_{ij}$) (for $i = j = d$) at state $(\mathbf{v},s)$ and the envelope condition (Env_i) (for $i = d$) at state $(\mathbf{w}_d,d)$ deliver $f_{dd}\big(\lambda_d(\mathbf{v},s) + \sum_{k=1}^{d-1}\mu_{dk}(\mathbf{v},s)\big) = f_{sd}P_d(\mathbf{w}_d,d) = f_{sd}\lambda_d(\mathbf{w}_d,d)$. Adding the common term $f_{sd}\sum_{k=1}^{d-1}\mu_{dk}(\mathbf{w}_d,d)$ to each side of the preceding equality then yields

$$\begin{aligned} f_{sd}\left[\lambda_d(\mathbf{w}_d,d) + \sum_{k=1}^{d-1}\mu_{dk}(\mathbf{w}_d,d)\right] &= f_{dd}\left[\lambda_d(\mathbf{v},s) + \sum_{k=1}^{d-1}\mu_{dk}(\mathbf{v},s)\right] \\ &\quad + f_{sd}\sum_{k=1}^{d-1}\mu_{dk}(\mathbf{w}_d,d). \end{aligned} \tag{C.4}$$

[7]The efficiency rays $\tilde{E}_i \subset \mathbb{R}^d_{++}$ are as described in Section 4.3, and the efficient sets $E_i \subset D$ are as described in Section 5.2 (where the assumption of CARA utility implies that each $E_i \subset D$ is a ray).

At state $(\mathbf{v}, s)$, summing the (FOCw$_{ij}$) for $i = d$ over $j \in S$ delivers

$$f_{sd} \mathsf{D_1} P(\mathbf{w}_d, d) = \lambda_d(\mathbf{v}, s) + \sum_{k=1}^{d-1} \mu_{dk}(\mathbf{v}, s). \tag{C.5}$$

At state $(\mathbf{w}_d, d)$, performing the analogous sum delivers

$$f_{dd} \mathsf{D_1} P(\tilde{\mathbf{w}}_d, d) = \lambda_d(\mathbf{w}_d, d) + \sum_{k=1}^{d-1} \mu_{dk}(\mathbf{w}_d, d). \tag{C.6}$$

Plugging (C.5) and (C.6) into (C.4) and dividing through by $f_{sd} \cdot f_{dd}$ yields (MS$_i$) for $i = d$.

Next, let $i < d$ be given. At state $(\mathbf{w}_d, d)$, summing the (FOCw$_{ij}$) over $j \in S$ yields

$$f_{di} \mathsf{D_1} P(\tilde{\mathbf{w}}_i, i) = \lambda_i(\mathbf{w}_d, d) + \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{w}_d, d) - \sum_{k=i+1}^{d} \mu_{ki}(\mathbf{w}_d, d). \tag{C.7}$$

Combining (C.6) and (C.7) delivers

$$\begin{aligned} \mathsf{D_1} P(\tilde{\mathbf{w}}_i, i) = \mathsf{D_1} P(\tilde{\mathbf{w}}_d, d) &- \frac{1}{f_{dd}} \sum_{k=1}^{d-1} \mu_{dk}(\mathbf{w}_d, d) - \frac{1}{f_{di}} \sum_{k=i+1}^{d} \mu_{ki}(\mathbf{w}_d, d) \\ &+ \frac{1}{f_{di}} \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{w}_d, d) - \left[\frac{\lambda_d(\mathbf{w}_d, d)}{f_{dd}} - \frac{\lambda_i(\mathbf{w}_d, d)}{f_{di}} \right]. \end{aligned} \tag{C.8}$$

The envelope conditions (Env$_i$) at state $(\mathbf{w}_d, d)$ imply that the final term in brackets equals $\frac{P_d(\mathbf{w}_d, d)}{f_{dd}} - \frac{P_i(\mathbf{w}_d, d)}{f_{di}}$, which vanishes by Lemma C.8 and ($\tilde{\mathrm{E}}_i$). Then, plugging the rest of (C.8) into the $i = d$ case of (MS$_i$) (established above) delivers (MS$_i$) for the given $i < d$. *Q.E.D.*

LEMMA C.10: *Given any* $(\mathbf{v}, s) \in D \times S$ *and* $i \in S$, *define* $\mathbf{w}_d := \xi^c(\mathbf{v}, s, d)$ *and* $\tilde{\mathbf{w}}_i := \xi^c(\mathbf{w}_d, d, i)$. *The following two properties are equivalent:*

(a) $\mathsf{D_1} P(\tilde{\mathbf{w}}_i, i) = \mathsf{D_1} P(\mathbf{w}_d, d)$ *for all* $i \in S$.

(b) $\mu_{ij}(\mathbf{w}_d, d) = 0$ *for all* $i, j \in S$.

PROOF: By (MS$_i$) in Lemma C.9, (b) implies (a). To show the converse, suppose that (a) holds. We proceed by induction through the type space. For the base step, let $i = d$. From (MS$_i$) for $i = d$ and dual feasibility (i.e., $\mu_{dk}(\cdot, \cdot) \geq 0$ on $D \times S$ for every $k \in S$), we see that $\mathsf{D_1} P(\tilde{\mathbf{w}}_d, d) = \mathsf{D_1} P(\mathbf{w}_d, d)$ implies that $\mu_{dk}(\mathbf{w}_d, d) = 0$ for all $k \in S$. For the inductive step, let $i < d$ be given and suppose that $\mu_{jk}(\mathbf{w}_d, d) = 0$ for all $j \geq i + 1$ and $k < j$. Then (MS$_i$) reduces to $\mathsf{D_1} P(\tilde{\mathbf{w}}_i, i) = \mathsf{D_1} P(\mathbf{w}_d, d) + (1/f_{di}) \sum_{k=1}^{i-1} \mu_{ik}(\mathbf{w}_d, d)$. As in the base step, it then follows from dual feasibility that $\mathsf{D_1} P(\tilde{\mathbf{w}}_i, i) = \mathsf{D_1} P(\mathbf{w}_d, d)$ implies that $\mu_{ik}(\mathbf{w}_d, d) = 0$ for all $k < i$. This completes the induction. We conclude that (a) implies (b), as desired. *Q.E.D.*

C.3.3. *An Equivalent "Interim" Formulation*

Herein, we introduce an equivalent *interim* formulation of the principal's problem, in which she optimizes over contractual variables contingent on the *current* period's report (rather than the previous period's report). This shift in timing convention is merely cosmetic, but useful for relating the Lagrange multipliers at a given $\mathbf{v} \in D$ across different $s \in S$.

Formally, given any $\mathbf{v} \in D$ and $i \in S$, we consider principal's i^{th} *interim problem*:

$$Q^i(\mathbf{v}) := \inf_{(u_i,\mathbf{w}_i)\in\mathcal{U}\times D} [C(u_i,i) + \alpha P(\mathbf{w}_i,i)] \tag{FE-Q^i}$$

$$\text{s.t.} \qquad v_i = u_i + \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i] \tag{PK$_i$}$$

$$v_j - v_i \geq \psi(u_i,j,i) - u_i + \alpha\left(\mathsf{E}^{\mathbf{f}_j}[\mathbf{w}_i] - \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i]\right) \tag{IC$^*_{ji}$}$$

for all $j \in S$ with $j > i$.[8] That is, if the agent reports to be of type $i \in S$ in the current period, the principal optimizes over flow and continuation utility pairs $(u_i,\mathbf{w}_i) \in \mathcal{U} \times D$, subject to promise keeping (PK$_i$) for type i and incentive compatibility (IC$^*_{ji}$) for all *higher* types $j > i$. Notably, the function $Q^i : D \to \mathbb{R}$ depends on $\mathbf{v}$ only through the components $(v_i, v_{i+1}, \ldots, v_d)$, as these are the only components that enter the constraints. For each $i \in S$, we define the i^{th} *constraint correspondence* $\Gamma_i : D \to \mathcal{U} \times D$ as

$$\Gamma_i(\mathbf{v}) := \left\{(u_i,\mathbf{w}_i) \in \mathcal{U} \times D : (u_i,\mathbf{w}_i) \text{ satisfies (PK}_i\text{) and (IC}^*_{ji}\text{) } \forall j \in S,\ j > i\right\}. \tag{C.9}$$

It is easy to see that, for every $\mathbf{v} \in D$, the constraint set $\Gamma(\mathbf{v})$ defined in (3.1) is given by the Cartesian product $\Gamma(\mathbf{v}) = \Gamma_1(\mathbf{v}) \times \cdots \times \Gamma_d(\mathbf{v})$.

LEMMA C.11: *For every $(\mathbf{v},s) \in D \times S$, the following properties hold:*
(a) *The value functions P and $\{Q^i\}_{i\in S}$ satisfy $P(\mathbf{v},s) = \sum_{i=1}^d f_{si} Q^i(\mathbf{v})$.*
(b) *A menu $(u_i,\mathbf{w}_i)_{i\in S} \in \Gamma(\mathbf{v})$ is a minimizer in (FE) at $(\mathbf{v},s)$ if and only if, for every $i \in S$, $(u_i,\mathbf{w}_i) \in \Gamma_i(\mathbf{v})$ is a minimizer in (FE-Q^i) at $\mathbf{v}$.*
Consequently, under the optimal contract, for every $(\mathbf{v},s) \in D \times S$ and $i \in S$, it holds that $\xi(\mathbf{v},s,i) = (\xi^f(\mathbf{v},s,i), \xi^c(\mathbf{v},s,i)) \in \Gamma_i(\mathbf{v})$ is a minimizer in (FE-Q^i) at $\mathbf{v}$.

PROOF: Because $\Gamma(\mathbf{v}) = \Gamma_1(\mathbf{v}) \times \cdots \times \Gamma_d(\mathbf{v})$, we can equivalently write the Bellman equation (FE) as $P(\mathbf{v},s) = \sum_{i=1}^d f_{si} \min_{(u_i,\mathbf{w}_i)\in\Gamma_i(\mathbf{v})} [C(u_i,i) + \alpha P(\mathbf{w}_i,i)] = \sum_{i=1}^d f_{si} Q^i(\mathbf{v})$. The lemma follows immediately from this observation and Proposition 3.2. *Q.E.D.*

COROLLARY C.12: *Under the optimal contract, for every $i \in S$, the function $\xi(\cdot,\cdot,i) : D \times S \to \mathcal{U} \times D$ depends on the argument $(\mathbf{v},s)$ only through the components $(v_i,\ldots,v_d)$.*

PROOF: Immediate from Lemma C.11 and the above observation that the constraint set $\Gamma_i(\mathbf{v})$ depends on $\mathbf{v}$ only through the components $(v_i,\ldots,v_d)$. *Q.E.D.*

LEMMA C.13: *For every $i \in S$, the following properties hold:*
(a) *The value function $Q^i : D \to \mathbb{R}$ is convex and continuously differentiable.*
(b) *For every $\mathbf{v} \in D$ and $i \in S$, there exists some $(u_i,\mathbf{w}_i) \in \Gamma_i(\mathbf{v})$ such that all of the (IC$^*_{ji}$) (for $j > i$) hold as strict inequalities.*

PROOF: For part (a), convexity follows from the definition of Q^i in (FE-Q^i) and the convexity of $P(\cdot,i)$ (Proposition 3.2). Since $P(\cdot,s) \in \mathbf{C}^1(D)$ (Proposition 3.2), Lemma C.11(a) and the sum rule for subdifferentials of convex functions imply that $\{\mathsf{D}P(\mathbf{v},s)\} = \partial P(\mathbf{v},s) =$

[8]Note that (IC$^*_{ji}$) is the same as (IC$^*_{ij}$) from the main text, except that the $i, j \in S$ indices are flipped.

$\sum_{i=1}^{d} f_{si}\partial Q^i(\mathbf{v})$ for every $(\mathbf{v},s) \in D \times S$.[9] Thus, for every $\mathbf{v} \in D$ and $i \in S$, $|\partial Q^i(\mathbf{v})| = 1$ and therefore Q^i is differentiable at $\mathbf{v}$. It follows that each $Q^i \in \mathbf{C}^1(D)$, as every differentiable convex function on the open set D (Theorem 3) is, in fact, continuously differentiable.

Part (b) follows from the fact that $\Gamma(\mathbf{v}) = \Gamma_1(\mathbf{v}) \times \cdots \times \Gamma_d(\mathbf{v})$ and Condition R.3. *Q.E.D.*

By Lemmas C.11 and C.13, the optimal $(u_i, \mathbf{w}_i) \in \Gamma_i(\mathbf{v})$ in (FE-Q^i) is characterized by saddle points of the Lagrangian (see, e.g., Exercise 7 on p. 236 and Theorem 2 on p. 221 of Luenberger (1969))

$$\begin{aligned}\mathscr{L}^i(\mathbf{v},\mathbf{u},\mathbf{w},\boldsymbol{\eta},\boldsymbol{\sigma}) &= C(u_i,i) + \alpha \sum_{k=1}^{d} f_{ik} Q^k(\mathbf{w}_i) + \eta_i\Big[v_i - u_i - \alpha\, \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i]\Big] \\ &\quad - \sum_{k=i+1}^{d} \sigma_{ki}\Big[v_k - v_i - \psi(u_i,k,i) + u_i - \alpha\big(\mathsf{E}^{\mathbf{f}_k}[\mathbf{w}_i] - \mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i]\big)\Big],\end{aligned} \tag{L$_i$}$$

where $\eta_i \in \mathbb{R}$ is the multiplier on (PK$_i$) and $\sigma_{ji} \geq 0$ is the multiplier on (IC$^*_{ji}$). We let $\eta_i(\mathbf{v}) \in \mathbb{R}$ denote the optimal multiplier on (PK$_i$) at $\mathbf{v} \in D$, and let $\sigma_{ji}(\mathbf{v}) \in \mathbb{R}_+$ denote the optimal multiplier on (IC$^*_{ji}$) at $\mathbf{v} \in D$. Lemma C.11 implies that, for every $(\mathbf{v},s) \in D \times S$ and $i \in S$, the interim problem (FE-Q^i) is (uniquely) solved by the pair $(u_i(\mathbf{v},s), \mathbf{w}_i(\mathbf{v},s))$ derived from the (unique) optimal menu $(u_i(\mathbf{v},s), \mathbf{w}_i(\mathbf{v},s))_{i\in S}$ characterized in SA–C.1 by the optimality conditions (Env$_i$), (FOCu_i), and (FOC$\mathbf{w}_{ij}$). We use this fact repeatedly below.

LEMMA C.14: *The following properties hold:*

(a) *For every* $(\mathbf{v},s) \in D \times S$ *and* $i \in S$, *the multipliers satisfy*

$$\frac{\lambda_i(\mathbf{v},s)}{f_{si}} = \eta_i(\mathbf{v}) + \sum_{k=i+1}^{d} \sigma_{ki}(\mathbf{v}) - \sum_{k=1}^{i-1} \frac{f_{sk}}{f_{si}} \sigma_{ik}(\mathbf{v}). \tag{C.10}$$

(b) *For every* $(\mathbf{v},s) \in D \times S$ *and* $i \in S$, *the multipliers satisfy*

$$\sum_{k=1}^{i-1}\Big[f_{sk}\sigma_{ik}(\mathbf{v}) - \mu_{ik}(\mathbf{v},s)\Big] = \sum_{k=i+1}^{d} \psi'(u_i(\mathbf{v},s),k,i)\Big[f_{si}\sigma_{ki}(\mathbf{v}) - \mu_{ki}(\mathbf{v},s)\Big]. \tag{C.11}$$

(c) *For every* $\mathbf{v} \in D$, *the following are equivalent: (i)* $\sigma_{ij}(\mathbf{v}) = 0$ *for all* $i, j \in S$, *(ii) for* some $s \in S$, $\mu_{ij}(\mathbf{v},s) = 0$ *for all* $i, j \in S$, *(iii) for* every $s \in S$, $\mu_{ij}(\mathbf{v},s) = 0$ *for all* $i, j \in S$.

PROOF: The proof proceeds by comparing the optimality conditions from SA–C.1 to those derived from the interim Lagrangians (L$_i$). For part (a), let $(\mathbf{v},s) \in D \times S$ be given. For each $i \in S$, the envelope conditions for (L$_i$) are

$$Q^i_j(\mathbf{v}) = \mathbf{1}(j = i) \cdot \Big[\eta_i(\mathbf{v}) + \sum_{k=i+1}^{d} \sigma_{ki}(\mathbf{v})\Big] - \mathbf{1}(j > i) \cdot \sigma_{ji}(\mathbf{v}) \tag{C.12}$$

[9] For any convex function $f : D \to \mathbb{R}$, $\partial f(\mathbf{v}) \subseteq \mathbb{R}^d$ denotes its subdifferential at $\mathbf{v} \in D$. Since D is open (Theorem 3), we have $\partial f(\mathbf{v}) \neq \varnothing$ for all $\mathbf{v} \in D$. See Chapters 23 and 25 of Rockafellar (1970) for the relevant facts about subdifferentials (in particular, Theorem 23.8, Theorem 25.1, and Corollary 25.5.1).

for all $j \in S$. Meanwhile, Lemma C.11(a) implies that $P_j(\mathbf{v}, s) = \sum_{i=1}^{d} f_{si} Q_j^i(\mathbf{v})$ for every $j \in S$. Thus, fixing $j \in S$ and summing over (C.12) over $i \in S$ delivers

$$P_j(\mathbf{v}, s) = f_{sj}\left[\eta_j(\mathbf{v}) + \sum_{k=j+1}^{d} \sigma_{kj}(\mathbf{v})\right] - \sum_{k=1}^{j-1} f_{sk}\sigma_{jk}(\mathbf{v}). \tag{C.13}$$

The envelope condition (Env$_i$) (for $i = j$) at $(\mathbf{v}, s)$ yields $\lambda_j(\mathbf{v}, s) = P_j(\mathbf{v}, s)$. Plugging this into (C.13) yields (C.10) (for $i = j$). Since the fixed $j \in S$ was arbitrary, this proves part (a).

For part (b), let $(\mathbf{v}, s) \in D \times S$ and $i \in S$ be given. The FOC for $u_i(\mathbf{v}, s)$ in (L$_i$) is

$$C'(u_i(\mathbf{v}, s), i) = \eta_i(\mathbf{v}) + \sum_{k=i+1}^{d} \sigma_{ki}(\mathbf{v}) \left(1 - \psi'(u_i(\mathbf{v}, s), k, i)\right). \tag{C.14}$$

Meanwhile, (FOCu_i) at $(\mathbf{v}, s)$ can be written as

$$C'(u_i(\mathbf{v}, s), i) = \frac{\lambda_i(\mathbf{v}, s)}{f_{si}} + \sum_{k=1}^{i-1} \frac{\mu_{ik}(\mathbf{v}, s)}{f_{si}} - \sum_{k=i+1}^{d} \psi'(u_i(\mathbf{v}, s), k, i)\frac{\mu_{ki}(\mathbf{v}, s)}{f_{si}}. \tag{C.15}$$

Equating (C.14) and (C.15) and substituting in (C.10) delivers

$$\begin{aligned}
&\frac{\lambda_i(\mathbf{v}, s)}{f_{si}} + \sum_{k=1}^{i-1} \frac{\mu_{ik}(\mathbf{v}, s)}{f_{si}} - \sum_{k=i+1}^{d} \psi'(u_i(\mathbf{v}, s), k, i)\frac{\mu_{ki}(\mathbf{v}, s)}{f_{si}} \\
&\quad = \sum_{k=i+1}^{d} \sigma_{ki}(\mathbf{v})\big(1 - \psi'(u_i(\mathbf{v}, s), k, i)\big) + \underbrace{\left[\frac{\lambda_i(\mathbf{v}, s)}{f_{si}} - \sum_{k=i+1}^{d} \sigma_{ki}(\mathbf{v}) + \sum_{k=1}^{i-1} \frac{f_{sk}}{f_{si}}\sigma_{ik}(\mathbf{v})\right]}_{= \eta_i(\mathbf{v})}.
\end{aligned}$$

Simplifying the above display yields (C.11), as desired.

For part (c), let $\mathbf{v} \in D$ be given. We show that (i) implies (iii) by induction. So suppose that (i) holds, and let $s \in S$ be given. For the base step, note that (C.11) with $i = d$ becomes $0 = \sum_{k=1}^{d-1} \left[\mu_{dk}(\mathbf{v}, s) - f_{sk}\sigma_{dk}(\mathbf{v})\right] = \sum_{k=1}^{d-1} \mu_{dk}(\mathbf{v}, s)$ where the second equality is because (i) holds at $\mathbf{v}$. Because $\mu_{dk}(\cdot, \cdot) \geq 0$ on $D \times S$ for all $k < d$, it follows that $\mu_{dk}(\mathbf{v}, s) = 0$ for all $k < d$. For the inductive step, let $\ell < d$ be given and suppose we have shown, for all $k > \ell$, that $\mu_{kj}(\mathbf{v}, s) = 0$ for all $j < k$. Then we have

$$\begin{aligned}
0 &= \sum_{k=1}^{\ell-1} \left[\mu_{\ell k}(\mathbf{v}, s) - f_{sk}\sigma_{\ell k}(\mathbf{v})\right] + \sum_{k=\ell+1}^{d} \psi'(u_\ell(\mathbf{v}, s), k, \ell)\left[f_{s\ell}\sigma_{k\ell}(\mathbf{v}) - \mu_{k\ell}(\mathbf{v}, s)\right] \\
&= \sum_{k=1}^{\ell-1} \mu_{\ell k}(\mathbf{v}, s) - \sum_{k=\ell+1}^{d} \psi'(u_\ell(\mathbf{v}, s), k, \ell)\mu_{k\ell}(\mathbf{v}, s) = \sum_{k=1}^{\ell-1} \mu_{\ell k}(\mathbf{v}, s),
\end{aligned}$$

where the first equality is (C.11) for $i = \ell$, the second equality is because (i) holds at $\mathbf{v}$, and the third equality is by the induction hypothesis. As before, it follows that $\mu_{\ell k}(\mathbf{v}, s) = 0$ for all $k < \ell$. This completes the induction. Since the given $s \in S$ was arbitrary, we conclude that (i) implies (iii). By full connectedness (Assumption Markov), the proof that (ii) implies (i) is completely analogous. Obviously, (iii) implies (ii). This proves part (c). *Q.E.D.*

LEMMA C.15: *For every* $\mathbf{v} \in D$*, the following hold:*

(a) *If there exists an* $s \in S$ *such that* $\mu_{ij}(\mathbf{v}, s) = 0$ *for all* $i, j \in S$*, then*

$$\frac{\lambda_i(\mathbf{v}, s')}{f_{s'i}} = \eta_i(\mathbf{v}) \qquad \text{for all } s', i \in S. \tag{C.16}$$

(b) *If there exists an* $s \in S$ *such that* $\mathbf{v} \in E_s$ *and* $\mu_{ij}(\mathbf{v}, s) = 0$ *for all* $i, j \in S$,[10] *then (i)* $\mathbf{v} \in E_{s'}$ *for all* $s' \in S$ *and (ii) there exists an* $\hat{\eta}(\mathbf{v}) \in \mathbb{R}$ *such that*

$$\frac{\lambda_i(\mathbf{v}, s')}{f_{s'i}} = \hat{\eta}(\mathbf{v}) \qquad \text{for all } s', i \in S. \tag{C.17}$$

PROOF: Let $\mathbf{v} \in D$ and such an $s \in S$ be given. For part (a), the hypothesis and Lemma C.14(c) imply that $\sigma_{ij}(\mathbf{v}) = 0$ for all $i, j \in S$. Plugging this into (C.10) (for the given $\mathbf{v}$ and across all $s', i \in S$) delivers (C.16). For part (b), the same logic delivers (C.16). At the same time, the hypothesis that $\mathbf{v} \in E_s$ and ($\tilde{\text{E}}_i$) (for $i = s$) imply that $\frac{P_1(\mathbf{v},s)}{f_{s1}} = \cdots = \frac{P_d(\mathbf{v},s)}{f_{sd}}$, which by the envelope condition (Env_i) at $(\mathbf{v}, s)$ is equivalent to $\frac{\lambda_1(\mathbf{v},s)}{f_{s1}} = \cdots = \frac{\lambda_d(\mathbf{v},s)}{f_{sd}}$. Plugging the latter into (C.16) delivers $\eta_1(\mathbf{v}) = \cdots = \eta_d(\mathbf{v})$. Denoting the common value of these multipliers by $\hat{\eta}(\mathbf{v}) \in \mathbb{R}$ yields (C.17). Now, let $s' \in S$ be given. We obtain from (C.17) that $\frac{\lambda_1(\mathbf{v},s')}{f_{s'1}} = \cdots = \frac{\lambda_d(\mathbf{v},s')}{f_{s'd}}$, which by the envelope condition (Env_i) at $(\mathbf{v}, s')$ is equivalent to $\frac{P_1(\mathbf{v},s')}{f_{s'1}} = \cdots = \frac{P_d(\mathbf{v},s')}{f_{s'd}}$. Then, ($\tilde{\text{E}}_i$) (for $i = s'$) and Lemma C.5 imply that $\mathbf{v} \in E_{s'}$, as desired. Since the given $s' \in S$ was arbitrary, this completes the proof of part (b). *Q.E.D.*

C.3.4. *First-Best Efficiency and Martingale Splitting*

The optimal contract ξ *self-generates at* $\mathbf{v} \in D$ if $\xi^c(\mathbf{v}, s, i) = \mathbf{v}$ for all $s, i \in S$. The optimal contract ξ is *first-best efficient at* $\mathbf{v} \in D$ if (i) ξ self-generates at $\mathbf{v}$, (ii) there is some $v \in \mathcal{U}$ such that $\mathbf{v} = v\mathbf{1}$, and (iii) $\xi^f(\mathbf{v}, s, i) = (1-\alpha)v$ for all $s, i \in S$.[11] First-best efficient contracts arise from solutions to the "first-best efficiency problem" (Eff_i^{FB}) defined in SA–F (see Lemma F.1(a) and Lemma F.2 therein), which is the analogue of the efficiency problem (Eff_i) without IC constraints.

LEMMA C.16: *The optimal contract* ξ *is not first-best efficient at any* $\mathbf{v} \in D$.

PROOF: If the domain D does not intersect the diagonal $\{\mathbf{v} \in \mathcal{U}^d : \exists v \in \mathcal{U} \text{ s.t. } \mathbf{v} = v\mathbf{1}\}$, then it is impossible for any $\mathbf{v} \in D$ to satisfy property (ii) in the definition of first-best efficiency, so we are done. Suppose, towards a contradiction, that D intersects the diagonal and that ξ is first-best efficient at some $\mathbf{v} \in D$, where $\mathbf{v} = v\mathbf{1}$ and $v \in \mathcal{U}$. Let $i, j \in S$ with $i > j$ be given. Then, for every $s \in S$, the (IC_{ij}^*) constraint at state $(\mathbf{v}, s)$ reads $v - v \geq \psi((1-\alpha)v, i, j) - (1-\alpha)v + \alpha\left(\mathsf{E}^{\mathbf{f}_i}[v\mathbf{1}] - \mathsf{E}^{\mathbf{f}_j}[v\mathbf{1}]\right)$, which reduces to $0 \geq \psi((1-\alpha)v, i, j) - (1-\alpha)v$. But $\psi(u, i, j) - u > 0$ for all $u \in \mathcal{U}$, a contradiction. *Q.E.D.*

LEMMA C.17: *Given any* $s \in S$ *and* $\mathbf{v} \in E_s$*, there exist some* $i, j \in S$ *(with* $i > j$*) such that* $\mu_{ij}(\mathbf{v}, s) > 0$.

[10] Recall the efficient sets $E_{s'} \subset D$ defined in Appendix C.3.1 (see, e.g., Lemma C.6(b)).

[11] Given condition (i), conditions (ii) and (iii) are equivalent (see, e.g., the proof of Lemma C.17 below).

PROOF: Let $s \in S$ and $\mathbf{v} \in E_s$ be given. Suppose, towards a contradiction, that $\mu_{ij}(\mathbf{v}, s) = 0$ for all $i, j \in S$. Then Lemma C.14(c) implies that $\mu_{ij}(\mathbf{v}, s') = 0$ for all $s', i, j \in S$ and Lemma C.15(b) implies that (C.17) holds. Plugging the former into the the optimality conditions (FOCu_i) and (FOC$\mathbf{w}_{ij}$) at $(\mathbf{v}, s')$ for every $s' \in S$ delivers

$$f_{s'i} C'(u_i(\mathbf{v}, s'), i) = \lambda_i(\mathbf{v}, s') \qquad \forall s', i \in S$$
$$f_{s'i} P_j(\mathbf{w}_i(\mathbf{v}, s'), i) = f_{ij} \lambda_i(\mathbf{v}, s') \qquad \forall s', i, j \in S.$$

Plugging (C.17) into the above display delivers

$$\frac{P_j(\mathbf{w}_i(\mathbf{v}, s'), i)}{f_{ij}} = \hat{\eta}(\mathbf{v}) = C'(u_i(\mathbf{v}, s'), i) \qquad \forall s', i, j \in S. \tag{C.18}$$

Plugging (C.17) into the envelope conditions (Env$_i$) at $(\mathbf{v}, i)$ for every $i \in S$ delivers

$$\frac{P_j(\mathbf{v}, i)}{f_{ij}} = \hat{\eta}(\mathbf{v}) \qquad \forall i, j \in S. \tag{C.19}$$

We claim that (C.18) and (C.19) together imply that the optimal contract ξ is first-best efficient at $\mathbf{v}$. Recall that, by definition, $\xi^f(\cdot, \cdot, i) = u_i(\cdot, \cdot)$ and $\xi^c(\cdot, \cdot, i) = \mathbf{w}_i(\cdot, \cdot)$ for all $i \in S$. To show that ξ self-generates at $\mathbf{v}$ (part (i) in the definition), note that combining (C.19) and the first equality in (C.18) delivers

$$\mathsf{D}P(\mathbf{v}, i) = \mathsf{D}P(\mathbf{w}_i(\mathbf{v}, s'), i) \qquad \forall s', i \in S.$$

It then follows from Lemma C.5 that $\mathbf{v} = \mathbf{w}_i(\mathbf{v}, s')$ for every $s', i \in S$, as desired. To establish parts (ii) and (iii) in the definition, first recall that $C(\cdot, i) := U^{-1}(\cdot) - \omega_i$ for every $i \in S$. Thus, the Inverse Function Theorem yields $C'(\cdot, i) = 1/U'(U^{-1}(\cdot))$ for every $i \in S$. Plugging this into the second equality in (C.18) delivers

$$U'(U^{-1}(u_i(\mathbf{v}, s'))) = \frac{1}{\hat{\eta}(\mathbf{v})} \qquad \forall s', i \in S.$$

Thus, since $U'(\cdot)$ and $U^{-1}(\cdot)$ are both injective (Assumption DARA), there exists some $z(\mathbf{v}) \in \mathcal{U}$ such that $u_i(\mathbf{v}, s') = z(\mathbf{v})$ for all $s', i \in S$. Given this property and the fact (shown above) that ξ self-generates at $\mathbf{v}$, the (PK$_i$) constraints (for all $i \in S$) at state $(\mathbf{v}, s')$ (for any $s' \in S$) yield $\mathbf{v} = z(\mathbf{v})\mathbf{1} + \alpha \mathbf{F} \mathbf{v}$, where $\mathbf{F} = [\mathbf{f}_i]_{i=1}^d$ is the matrix of transition probabilities. Thus, $\mathbf{v} = z(\mathbf{v})(\mathbf{I} - \alpha \mathbf{F})^{-1}\mathbf{1} = \frac{z(\mathbf{v})}{1-\alpha}\mathbf{1}$.[12] That is, $\mathbf{v} = v\mathbf{1}$ for $v := \frac{z(\mathbf{v})}{1-\alpha} \in \mathcal{U}$, and hence $u_i(\mathbf{v}, s') = (1 - \alpha)v$ for all $s', i \in S$. This establishes that ξ is first-best efficient at $\mathbf{v}$.

Since Lemma C.16 establishes that ξ cannot be first-best efficient at $\mathbf{v}$, this delivers the desired contradiction. We conclude that $\mu_{ij}(\mathbf{v}, s) > 0$ for some $i, j \in S$ (with $i > j$). *Q.E.D.*

LEMMA C.18: *Given any $(\mathbf{v}, s) \in D \times S$, define $\mathbf{w}_d := \xi^c(\mathbf{v}, s, d)$ and $\tilde{\mathbf{w}}_i := \xi^c(\mathbf{w}_d, d, i)$ for every $i \in S$. There exists an $i \in S$ such that $\mathsf{D}_1 P(\tilde{\mathbf{w}}_i, i) \neq \mathsf{D}_1 P(\mathbf{w}_d, d)$.*

[12] In particular, note that $\mathbf{v} = z(\mathbf{v})\mathbf{1} + \alpha\mathbf{F}\mathbf{v}$ can be rewritten as $(\mathbf{I} - \alpha\mathbf{F})\mathbf{v} = z(\mathbf{v})\mathbf{1}$, where $\mathbf{I} \in \mathbb{R}^{d \times d}$ is the identity matrix. The matrix $\mathbf{I} - \alpha\mathbf{F}$ is strictly diagonally dominant: for every $i \in S$, $|[\mathbf{I} - \alpha\mathbf{F}]_{ii}| = 1 - \alpha f_{ii} > \alpha(1 - f_{ii}) = \sum_{j \neq i} |[\mathbf{I} - \alpha\mathbf{F}]_{ij}|$. Thus, $\mathbf{I} - \alpha\mathbf{F}$ is invertible. Since $(\mathbf{I} - \alpha\mathbf{F})\mathbf{1} = (1 - \alpha)\mathbf{1}$, we have $(\mathbf{I} - \alpha\mathbf{F})^{-1}\mathbf{1} = \frac{1}{1-\alpha}\mathbf{1}$. It follows that $\mathbf{v} = z(\mathbf{v})(\mathbf{I} - \alpha\mathbf{F})^{-1}\mathbf{1} = \frac{z(\mathbf{v})}{1-\alpha}\mathbf{1}$, as claimed.

PROOF: By Lemma C.8, we have $\mathbf{w}_d \in E_d$. Thus, Lemma C.17 implies that there exist some $i, j \in S$ (with $i > j$) such that $\mu_{ij}(\mathbf{w}_d, d) > 0$. Lemma C.10 then implies that there exists an $i \in S$ such that $\mathsf{D}_1 P(\tilde{\mathbf{w}}_i, i) \neq \mathsf{D}_1 P(\mathbf{w}_d, d)$, as desired. *Q.E.D.*

The main implication of Lemma C.18 is that, along the optimal path, the marginal cost martingale is non-constant with strictly positive probability at every state of the form $(\mathbf{w}_d, d)$. That is, the martingale "splits" with strictly positive probability at such histories.

C.4. *Main Proof of Theorem 1*

We use the following notation throughout the proof. Recall from Section 3 that the space of *paths* of type reports is $\mathscr{H} := S^\infty$ with generic element $h := \left(s^{t+1}\right)_{t=0}^{\infty}$, where s^t denotes the (truthfully) reported type in period $t-1$. Let $\tau^{(t)}$ denote the *random* time defined pathwise by $\tau^{(t)}(h) := \sup\left\{T \leq t : \ s^T = d\right\}$. That is, given path h, $\tau^{(t)}(h)$ is the last date (i) that precedes or equals t and (ii) that was immediately preceded by a realized endowment ω_d. In particular, the state in period $\tau^{(t)}$ is $(\mathbf{v}, d)$ for some $\mathbf{v} \in E_d$ (by Lemma C.8), where E_d is defined in Appendix C.3.1. It is easy to see that $\tau^{(t)}$ is a well-defined stopping time, that the process $\left(\tau^{(t)}\right)_{t=0}^{\infty}$ is non-decreasing, and that $\lim_{t\to\infty} \tau^{(t)} = \infty$, P-a.s.

Martingale Convergence. Define the events $\mathscr{I}, \mathscr{H}^*, \mathscr{I}^* \subseteq \mathscr{H}$ as follows:

$$\mathscr{I} := \left\{h \in \mathscr{H} : \forall i \in S,\ (s^t, s^{t+1}) = (d, i) \text{ occurs for infinitely many } t\right\},$$

$$\mathscr{H}^* := \left\{h \in \mathscr{H} : \lim_{t\to\infty} \mathsf{D}_1 P(\mathbf{v}^{(t)}(h), s^t) \text{ exists}\right\}, \text{ and } \mathscr{I}^* := \mathscr{I} \cap \mathscr{H}^*.$$

We will repeatedly make use of the following facts. First, $\mathscr{I}$ satisfies two properties: (i) $\lim_{t\to\infty} \tau^{(t)}(h) = +\infty$ for all $h \in \mathscr{I}$ by construction, and (ii) $\mathsf{P}(\mathscr{I}) = 1$ by standard facts about Markov chains (see Corollary G.3 in SA-G). Second, $\mathscr{H}^*$ satisfies $\mathsf{P}(\mathscr{H}^*) = 1$ by Doob's Martingale Convergence Theorem (see Theorem 2 in Shiryaev (1995, p. 517)), because the process $\left(\mathsf{D}_1 P(\mathbf{v}^{(t)}, s^{(t)})\right)_{t=0}^{\infty}$ is a strictly positive martingale (Proposition 4.4). Finally, $\mathscr{I}^*$ satisfies $\mathsf{P}(\mathscr{I}^*) = 1$, being the finite intersection of full-measure events.

LEMMA C.19: *The marginal cost martingale* $\mathsf{D}_1 P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ *almost surely.*

PROOF: It suffices to show that $\mathsf{D}_1 P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ on the event $\mathscr{I}^*$. So, fix a path $h = \left(s^{t+1}\right)_{t=0}^{\infty} \in \mathscr{I}^*$. Since the path is fixed, let $\tau^t := \tau^{(t)}(h)$ and $\mathbf{v}^t := \mathbf{v}^{(t)}(h)$ for all $t \in \mathbb{N}$. For all $i \in S$ and $t \in \mathbb{N}$, define $\tau_i^t := \sup\left\{T \leq t : (s^T, s^{T+1}) = (d, i)\right\}$. By construction, $s^{\tau_i^t + 1} = i$ for all $i \in S$ and $t \in \mathbb{N}$, and $\lim_{t\to\infty} \tau_i^t = +\infty$ for all $i \in S$.

Suppose, towards a contradiction, that $\mathsf{D}_1 P(\mathbf{v}^t, s^t) \to C > 0$. By construction, $\mathbf{v}^{\tau^t} = \xi^c(\mathbf{v}^{\tau^t - 1}, s^{\tau^t - 1}, d)$ and $s^{\tau^t} = d$. Thus, Lemma C.8 and ($\tilde{\mathrm{E}}_i$) imply that $\mathbf{v}^{\tau^t} \in E_d$ and $\mathsf{D}P(\mathbf{v}^{\tau^t}, s^{\tau^t} = d) \in \tilde{E}_d$ for all $t \in \mathbb{N}$. This and the supposition then imply that $\mathsf{D}P(\mathbf{v}^{\tau^t}, s^{\tau^t} = d) \to \mathbf{y}^* := (Cf_{d1}, \dots, Cf_{dd}) \in \tilde{E}_d$. Since $\mathsf{D}P(\cdot, d)$ is a homeomorphism (Lemma C.5), it follows that $\mathbf{v}^{\tau^t} \to \mathbf{w}_d^* := [\mathsf{D}P(\cdot, d)]^{-1}(\mathbf{y}^*) \in E_d$. Thus, for every $i \in S$, we have $\xi^c(\mathbf{v}^{\tau^t}, d, i) \to \xi^c(\mathbf{w}_d^*, d, i) =: \tilde{\mathbf{w}}_i^*$ because the policy functions are continuous (Proposition 3.2). Since $\mathbf{v}^{\tau_i^t + 1} = \xi^c(\mathbf{v}^{\tau_i^t}, s^{\tau_i^t} = d, s^{\tau_i^t + 1} = i)$ by construction, it follows that $\lim_{t\to\infty} \mathsf{D}_1 P(\mathbf{v}^{\tau_i^t + 1}, s^{\tau_i^t + 1} = i) = \mathsf{D}_1 P(\tilde{\mathbf{w}}_i^*, i)$ for each $i \in S$, because each $P(\cdot, i)$ is continuously differentiable (Proposition 3.2). By the supposition, it follows that $\mathsf{D}_1 P(\mathbf{w}_d^*, d) = C = \mathsf{D}_1 P(\tilde{\mathbf{w}}_i^*, i)$ for all $i \in S$. But this violates Lemma C.18, yielding the desired contradiction. Thus, $\mathsf{D}_1 P(\mathbf{v}^t, s^t) \to 0$, as desired. *Q.E.D.*

Convergence of Multipliers. We next establish convergence of the Lagrange multipliers. For every state $(\mathbf{v}, s) \in D \times S$, define the following vectors. Denote by $\boldsymbol{\lambda}(\mathbf{v}, s) := \big(\lambda_1(\mathbf{v}, s), \ldots, \lambda_d(\mathbf{v}, s)\big) \in \mathbb{R}^d$ the vector of multipliers on the (PK$_i$) constraints. For each $j \in S$, denote by $\boldsymbol{\mu}_{*,j}(\mathbf{v}, s) := \big(\mu_{j+1,j}(\mathbf{v}, s), \ldots, \mu_{d,j}(\mathbf{v}, s)\big) \in \mathbb{R}_+^{d-j}$ the vector of multipliers on the (IC$_{ij}$) constraints (for all $i > j$). Let $\boldsymbol{\mu}(\mathbf{v}, s) := \big(\boldsymbol{\mu}_{*,j}(\mathbf{v}, s)\big)_{j \in S} \in \mathbb{R}^{d(d-1)/2}$ denote the vector that stacks the $\boldsymbol{\mu}_{*,j}(\mathbf{v}, s)$. Finally, let $\boldsymbol{\nu}(\mathbf{v}, s) := (\boldsymbol{\lambda}(\mathbf{v}, s), \boldsymbol{\mu}(\mathbf{v}, s)) \in \mathbb{R}^{d(d+1)/2}$ denote the vector that stacks all of the multipliers. We proceed through a series of lemmas.

LEMMA C.20: *It holds that $\boldsymbol{\nu}^{(\tau^{(t)})} \to \mathbf{0}$ and $\mathsf{D}P(\tilde{\mathbf{w}}_i^{(\tau^{(t)})}, i) \to \mathbf{0}$ for all $i \in S$ almost surely, where $\tilde{\mathbf{w}}_i^{(\tau^{(t)})} := \xi^c(\mathbf{v}^{(\tau^{(t)})}, s^{(\tau^{(t)})} = d, i)$.*

PROOF: Again, it suffices to show convergence on $\mathcal{J}^*$. So, fix a path $h = \big(s^{t+1}\big)_{t=0}^{\infty} \in \mathcal{J}^*$ and let $\tau^t := \tau^{(t)}(h)$ and $\mathbf{v}^t := \mathbf{v}^{(t)}(h)$. As shown in the proof of Lemma C.19, we have $\mathsf{D}_1 P(\mathbf{v}^t, s^t) \to 0$ along this path. Since $\mathbf{v}^{\tau^t} = \xi^c(\mathbf{v}^{\tau^t - 1}, s^{\tau^t - 1}, d)$ and $s^{\tau^t} = d$ by construction, Lemma C.8 and ($\tilde{\text{E}}_i$) further imply that $\mathsf{D}P(\mathbf{v}^{\tau^t}, s^{\tau^t} = d) \to \mathbf{0}$. Then (Env$_i$) implies that

$$\boldsymbol{\lambda}^{\tau^t} := \boldsymbol{\lambda}(\mathbf{v}^{\tau^t}, s^{\tau^t} = d) \to \mathbf{0}. \tag{C.20}$$

We next show that $\boldsymbol{\mu}_{*,i}^{\tau^t} \to \mathbf{0}$ and $\mathsf{D}P(\tilde{\mathbf{w}}^{\tau_i^t}, i) \to \mathbf{0}$ for all $i \in S$. To do so, we proceed by induction through the type space, starting from the bottom.

Base step: The first-order condition (FOC$\mathbf{w}_{ij}$) with $i = 1$ at state $(\mathbf{v}^{\tau^t}, d)$ is $f_{d1} P_j(\tilde{\mathbf{w}}_1^{\tau^t}, 1) = f_{1j}\big(\lambda_1(\mathbf{v}^{\tau^t}, d) + 0\big) - \sum_{k=2}^{d} f_{kj}\mu_{k1}(\mathbf{v}^{\tau^t}, d)$. Because $\boldsymbol{\mu}(\mathbf{v}^{\tau^t}, d) \geq \mathbf{0}$ (dual feasibility), it follows from (C.20) and full connectedness (Assumption Markov) that $\limsup_{t\to\infty} P_j(\tilde{\mathbf{w}}_1^{\tau^t}, 1) \leq 0$ for all $j \in S$. But since $\mathsf{D}_1 P(\tilde{\mathbf{w}}_1^{\tau^t}, 1) \geq 0$ for all $t \in \mathbb{N}$ by Lemma C.1, we obtain $P_j(\tilde{\mathbf{w}}_1^{\tau^t}, 1) \to 0$ for all $j \in S$. It then follows from the FOC that $\sum_{k=2}^{d} f_{kj}\mu_{k1}(\mathbf{v}^{\tau^t}, d) \to 0$ for all $j \in S$, and hence $\mu_{k1}(\mathbf{v}^{\tau^t}, d) \to 0$ for all $2 \leq k \leq d$. Putting this together, we obtain

$$\mathsf{D}P(\tilde{\mathbf{w}}_1^{\tau^t}, 1) \to \mathbf{0} \quad \text{and} \quad \boldsymbol{\mu}_{*,1}^{\tau^t} := \boldsymbol{\mu}_{*,1}(\mathbf{v}^{\tau^t}, d) \to \mathbf{0}. \tag{C.21}$$

Inductive step: Let $2 \leq m \leq d$. Suppose we have shown that $\mathsf{D}P(\tilde{\mathbf{w}}_k^{\tau^t}, k) \to \mathbf{0}$ and $\boldsymbol{\mu}_{*,k}(\mathbf{v}^{\tau^t}, d) \to \mathbf{0}$ for all $1 \leq k < m$. The first order condition (FOC$\mathbf{w}_{ij}$) with $i = m$ at state $(\mathbf{v}^{\tau^t}, d)$ is $f_{dm} P_j(\tilde{\mathbf{w}}_m^{\tau^t}, m) = f_{mj}\big(\lambda_m(\mathbf{v}^{\tau^t}, d) + \sum_{k=1}^{m-1}\mu_{mk}(\mathbf{v}^{\tau^t}, d)\big) - \sum_{k=m+1}^{d} f_{kj}\mu_{km}(\mathbf{v}^{\tau^t}, d)$. The inductive hypothesis and (C.20) imply that $\lambda_m(\mathbf{v}^{\tau^t}, d) + \sum_{k=1}^{m-1}\mu_{mk}(\mathbf{v}^{\tau^t}, d) \to \mathbf{0}$. As in the base step, invoking dual feasibility, full connectedness, and Lemma C.1 then implies that $P_j(\tilde{\mathbf{w}}_m^{\tau^t}, m) \to 0$ and $\mu_{km}(\mathbf{v}^{\tau^t}, d) \to 0$ for all $m + 1 \leq k \leq d$. We conclude that

$$\mathsf{D}P(\tilde{\mathbf{w}}_m^{\tau^t}, m) \to \mathbf{0} \quad \text{and} \quad \boldsymbol{\mu}_{*,m}^{\tau^t} := \boldsymbol{\mu}_{*,m}(\mathbf{v}^{\tau^t}, d) \to \mathbf{0}. \tag{C.22}$$

This completes the induction. Combining (C.20), (C.21), and (C.22) yields the lemma. *Q.E.D.*

LEMMA C.21: *For every $k \in \mathbb{N} \cup \{0\}$, $\boldsymbol{\nu}^{(\tau^{(t)}+k)} \to \mathbf{0}$ and $\mathsf{D}P(\tilde{\mathbf{w}}_i^{(\tau^{(t)}+k)}, i) \to \mathbf{0}$ as $t \to \infty$ for all $i \in S$ almost surely, where $\tilde{\mathbf{w}}_i^{(\tau^{(t)}+k)} := \xi^c(\mathbf{v}^{(\tau^{(t)}+k)}, s^{(\tau^{(t)}+k)}, i)$.*

PROOF: Again, it suffices to show convergence on $\mathcal{J}^*$. So, fix a path $h = \big(s^{t+1}\big)_{t=0}^{\infty} \in \mathcal{J}^*$ and let $\tau^t := \tau^{(t)}(h)$ and $\mathbf{v}^t := \mathbf{v}^{(t)}(h)$. We proceed by induction on k, with Lemma C.20 serving as the base $(k = 0)$ step.

For the inductive step, let $k \in \mathbb{N}$ be given. Suppose we have shown that, for every $0 \leq m < k$, (i) $\lim_{t\to\infty} \boldsymbol{\nu}^{\tau^t+m} = \mathbf{0}$ and (ii) $\lim_{t\to\infty} \mathsf{D}P(\tilde{\mathbf{w}}_i^{\tau^t+m}, i) = \mathbf{0}$ for all $i \in S$. Since $\mathbf{v}^{\tau^t+k} = \xi^c(\mathbf{v}^{\tau^t+k-1}, s^{\tau^t+k-1}, s^{\tau^t+k})$ by construction, we obtain that $\mathbf{v}^{\tau^t+k} = \tilde{\mathbf{w}}_{s^{\tau^t+k}}^{\tau^t+k-1}$ from the definition of $(\tilde{\mathbf{w}}_i^{\tau^t+k-1})_{i\in S}$. Thus, since S is a finite set, part (ii) of the inductive hypothesis yields $\lim_{t\to\infty} \mathsf{D}P(\mathbf{v}^{\tau^t+k}, s^{\tau^t+k}) = \mathbf{0}$. The envelope condition (Env$_i$) then yields

$$\boldsymbol{\lambda}^{\tau^t+k} := \boldsymbol{\lambda}(\mathbf{v}^{\tau^t+k}, s^{\tau^t+k}) \to \mathbf{0} \quad \text{as } t \to \infty. \tag{C.23}$$

We can then show that

$$\mathsf{D}P(\tilde{\mathbf{w}}_i^{\tau^t+k}, i) \to \mathbf{0} \quad \text{and} \quad \boldsymbol{\mu}_{*,i}^{\tau^t+k} := \boldsymbol{\mu}_{*,i}(\mathbf{v}^{\tau^t+k}, s^{\tau^t+k}) \to \mathbf{0} \quad \forall i \in S \quad \text{as} \quad t \to \infty \tag{C.24}$$

by replicating the "induction through the type space" argument from the proof of Lemma C.20, with two minor modifications: (a) replace the date τ^t with the new date $\tau^t + k$ everywhere the former appears, and (b) replace the state $(\mathbf{v}^{\tau^t}, s^{\tau^t} = d)$ with the new state $(\mathbf{v}^{\tau^t+k}, s^{\tau^t+k})$ wherever the former appears. The details are straightforward, and thus omitted.

Combining (C.23) and (C.24) completes the induction on $k \in \mathbb{N}$, and the proof. *Q.E.D.*

LEMMA C.22: *The vector of multipliers $\boldsymbol{\nu}^{(t)} \to \mathbf{0}$ in probability.*

PROOF: Define the stochastic processes $\left(\delta^{(t)}\right)_{t=0}^{\infty}$ and $\left(L^{(t)}\right)_{t=0}^{\infty}$ by $\delta^{(t)} := \|\boldsymbol{\nu}^{(t)}\|$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^{d(d+1)/2}$, and $L^{(t)} := t - \tau^{(t)}$. Thus, along path $h \in \mathscr{H}$, $\delta^{(t)}(h) \geq 0$ denotes the distance of $\boldsymbol{\nu}^{(t)}(h)$ from the zero vector and $L^{(t)}(h) \in \mathbb{N} \cup \{0\}$ denotes the lag since the last time that $s = d$.

To show convergence in probability, we must show that $\lim_{t\to\infty} \mathbf{P}(\delta^{(t)} > \varepsilon) = 0$ for all $\varepsilon > 0$. So, let $\varepsilon > 0$ be given. For every $t, k \in \mathbb{N}$, define the events $A_{\varepsilon,t} := \left\{h \in \mathscr{H} : \delta^{(t)}(h) > \varepsilon\right\}$, $B_{k,t} := \left\{h \in \mathscr{H} : L^{(t)}(h) > k\right\}$, and $C_{\varepsilon,k,t} := \bigcup_{T\geq t} \left[A_{\varepsilon,T} \cap B_{k,T}^c\right]$. For every $t, k \in \mathbb{N}$, we have $C_{\varepsilon,k,t+1} \subseteq C_{\varepsilon,k,t}$ and $A_{\varepsilon,t} \cap B_{k,t}^c \subseteq C_{\varepsilon,k,t}$. Consequently, we have:

$$\forall k, t \in \mathbb{N}, \quad \mathbf{P}(A_{\varepsilon,t}) = \mathbf{P}\left(A_{\varepsilon,t} \cap B_{k,t}^c\right) + \mathbf{P}(A_{\varepsilon,t} \cap B_{k,t}) \leq \mathbf{P}(C_{\varepsilon,k,t}) + \mathbf{P}(B_{k,t}). \tag{C.25}$$

To complete the proof, we must show that $\lim_{t\to\infty} \mathbf{P}(A_{\varepsilon,t}) = 0$. We do this in two steps, corresponding to the two terms on the RHS of (C.25).

Step 1: We claim that $\lim_{t\to\infty} \mathbf{P}(C_{\varepsilon,k,t}) = 0$ for every $k \in \mathbb{N}$. In particular, we have

$$\begin{aligned}\lim_{t\to\infty} \mathbf{P}(C_{\varepsilon,k,t}) &= \mathbf{P}(\cap_{t\in\mathbb{N}} C_{\varepsilon,k,t}) \\ &= \mathbf{P}\left(\left\{h \in \mathscr{H} : \delta^{(t)}(h) > \varepsilon \text{ and } L^{(t)}(h) \leq k \text{ for infinitely many } t\right\}\right) = 0,\end{aligned}$$

where the first equality is by continuity of probability (as the sequence $(C_{\varepsilon,k,t})_{t\in\mathbb{N}}$ is decreasing), the second equality is by definition, and the third equality is by the definition of $L^{(t)}$ and Lemma C.21. This proves the claim.

Step 2: Together, (C.25) and Step 1 deliver the following fact:

$$\forall k \in \mathbb{N}, \quad \limsup_{t\to\infty} \mathbf{P}(A_{\varepsilon,t}) \leq \limsup_{t\to\infty} \mathbf{P}(B_{k,t}). \tag{C.26}$$

Define the map $H : \mathbb{N} \to [0, 1]$ by $H(k) := 1 - \limsup_{t\to\infty} \mathbf{P}(B_{k,t})$. Note that H is non-decreasing (since $B_{k+1,t} \subseteq B_{k,t}$ for every $k, t \in \mathbb{N}$). Consequently, if we can show that

$\lim_{k\to\infty} H(k) = 1$, we are done. In particular, since (C.26) holds for all k and only its RHS depends on k, it will then follow that $\limsup_{t\to\infty} \mathbf{P}(A_{\varepsilon,t}) \le \lim_{k\to\infty}(1 - H(k)) = 0$, and hence that $\lim_{t\to\infty} \mathbf{P}(A_{\varepsilon,t}) = 0$ (since probability is non-negative), as desired. Thus, the remainder of the proof is dedicated to showing that, in fact, $\lim_{k\to\infty} H(k) = 1$.

To this end, we appeal to facts about Markov chains. Under Assumption Markov, the type process $\left(s^{(t)}\right)$ is ergodic; in particular, there exists a unique stationary distribution $\boldsymbol{\pi} \in \Delta(S)$ such that $\lim_{t\to\infty} \mathbf{P}(s^{(t)} = i) = \pi_i > 0$ for all $i \in S$. Denote by $\left(r^{(t)}\right)$ the *time-reversed* type process, i.e., the S-valued Markov chain with transition probabilities $\mathbf{Q}\left(r^{(t+1)} = j \;\middle|\; r^{(t)} = i\right) = g_{ij} := \frac{\pi_j}{\pi_i} \cdot f_{ji}$ and induced measure over paths $\mathbf{Q} \in \Delta(S^\infty)$.[13] Denote by T_d^R the first hitting time of state d for the time-reversed chain, i.e., the random variable $T_d^R := \inf\left\{t \in \mathbb{N} \cup \{0\} : r^{(t)} = d\right\}$. For each $i \in S$, define the map $H_i : \mathbb{N} \to [0,1]$ as $H_i(k) := \mathbf{Q}\left(T_d^R \le k \;\middle|\; r^{(0)} = i\right)$. We claim that the following three properties hold:

(a) For every $i \in S$, $\lim_{k\to\infty} H_i(k) = 1$.
(b) For every $i \in S$ and $k \in \mathbb{N}$, $H_i(k) = 1 - \lim_{t\to\infty} \mathbf{P}\left(B_{k,t} \mid s^{(t)} = i\right)$.
(c) For every $k \in \mathbb{N}$, $H(k) = \sum_{i\in S} \pi_i H_i(k)$.

Together, (a)–(c) imply that $\lim_{k\to\infty} H(k) = 1$, so establishing (a)–(c) completes the proof.

For property (a), note that $\lim_{k\to\infty} H_i(k) = \mathbf{Q}\left(T_d^R < \infty \;\middle|\; r^{(0)} = i\right)$ for all $i \in S$. Furthermore, $\mathbf{Q}\left(T_d^R < \infty \;\middle|\; r^{(0)} = i\right) = 1$ for all $i \in S$ because the time-reversed chain is fully connected, and hence all states communicate with each other and are recurrent. This establishes property (a).

For property (b), let $i \in S$ and $k \in \mathbb{N}$ be given. Since $\lim_{t\to\infty} \Pr(s^{(t)} = i) = \pi_i > 0$, there exists $T \in \mathbb{N}$ such that $\Pr(s^{(t)} = i) > 0$ for all $t \ge T$. Then for $t \ge T$, we have

$$\mathbf{P}\left(B_{k,t} \mid s^{(t)} = i\right) = \frac{\mathbf{P}\left(s^{(t-m)} \ne d \ \forall\, m = 0,\ldots,k \text{ and } s^{(t)} = i\right)}{\mathbf{P}\left(s^{(t)} = i\right)}$$

by definition of $B_{k,t}$ and conditional probability. For $i = d$, we have $\mathbf{P}(B_{k,t} \mid s^{(t)} = i) = 0 = 1 - H_i(k)$ for all $t \ge T$, so we are done. For any $i \ne d$, we have

$$\begin{aligned}
&\mathbf{P}\left(s^{(t-m)} \ne d \ \forall\, m = 0,\ldots,k \text{ and } s^{(t)} = i\right) \\
&\quad = \sum_{(j_1,\ldots,j_k)\in\{1,\ldots,d-1\}^k} \mathbf{P}\left(s^{(t-k)} = j_k\right) \cdot f_{j_k,j_{k-1}} \cdots f_{j_2,j_1} \cdot f_{j_1,i} \\
&\quad = \sum_{(j_1,\ldots,j_k)\in\{1,\ldots,d-1\}^k} \pi_i \cdot \left(g_{i,j_1} \cdots g_{j_{k-1},j_k}\right) \cdot \frac{\mathbf{P}\left(s^{(t-k)} = j_k\right)}{\pi_{j_k}}
\end{aligned}$$

where the first equality uses the Markov property (of the forward chain) to sum over all paths from potential starting points $s^{(t-k)} \in S$ to the endpoint $s^{(t)} = i$, and the second equality is by definition of the time-reversed transition probabilities. Consequently,

$$\lim_{t\to\infty} \mathbf{P}\left(B_{k,t} \mid s^{(t)} = i\right) = \sum_{(j_1,\ldots,j_k)\in\{1,\ldots,d-1\}^k} \left(g_{i,j_1} \cdots g_{j_{k-1},j_k}\right) = \mathbf{Q}\left(T_d^R > k \mid r^{(0)} = i\right)$$

[13] The backward transition probabilities g_{ij} are defined from the forward transition probabilities f_{ij} via Bayes' Rule, with the stationary distribution of the forward chain, $\boldsymbol{\pi}$, serving as the "prior."

where the first equality is by the preceding two displays and ergodicity of $(s^{(t)})$, and the second equality is by definition of T_d^R. Since $H_i(k) = 1 - \mathbf{Q}\left(T_d^R > k \mid r^{(0)} = i\right)$ by construction, this establishes property (b).

Finally, for property (c), note that $\mathbf{P}(B_{k,t}) = \sum_{i\in S} \mathbf{P}(s^{(t)} = i)\, \mathbf{P}(B_{k,t} \mid s^{(t)} = i)$. Ergodicity of $(s^{(t)})$ and property (b) imply that $1 - H(k) = \lim_{t\to\infty} \mathbf{P}(B_{k,t}) = \sum_{i\in S} \lim_{t\to\infty} \mathbf{P}(s^{(t)} = i) \cdot \lim_{t\to\infty} \mathbf{P}(B_{k,t} \mid s^{(t)} = i) = \sum_{i\in S} \pi_i\,(1 - H_i(k)) = 1 - \sum_{i\in S} \pi_i H_i(k)$, as desired. *Q.E.D.*

Convergence of Allocations. Our final lemma shows that convergence of the multipliers implies convergence of the allocation. Recall that $u_i(\mathbf{v}, s) = \xi^f(\mathbf{v}, s, i)$ is the flow utility given to a type-i agent in state $(\mathbf{v}, s)$. To ease notation, for each $i \in S$, define the process $(u_i^{(t)})_{t=0}^{\infty}$ by $u_i^{(t)} := u_i(\mathbf{v}^{(t)}, s^{(t)})$.

LEMMA C.23: *Under the optimal contract, $u_i^{(t)} \to -\infty$ in probability for all $i \in S$.*

PROOF: Let $i \in S$ be given. The first-order condition (FOCu_i) at state $(\mathbf{v}^{(t)}, s^{(t)})$ is

$$f_{s^{(t)}i} C'\left(u_i^{(t)}, i\right) = \underbrace{\lambda_i^{(t)} + \sum_{k=1}^{i-1} \mu_{ik}^{(t)}}_{=:\, A_i^{(t)}} - \underbrace{\sum_{k=i+1}^{d} \psi'\left(u_i^{(t)}, k, i\right) \mu_{ki}^{(t)}}_{=:\, B_i^{(t)}}$$

Note that $B_i^{(t)} \geq 0$ because $\mu_{ki}^{(t)} \geq 0$ and $\psi'(\cdot, k, i) > 0$ for all $k, i \in S$. Since $C'(\cdot, i) > 0$, it follows that $0 < C'(u_i^{(t)}, i) \leq A_i^{(t)}/f_{s^{(t)}i}$. Lemma C.22 and full connectedness (Assumption Markov) imply that $A_i^{(t)}/f_{s^{(t)}i} \to 0$ in probability. Thus, we obtain $C'(u_i^{(t)}, i) \to 0$ in probability. Finally, because $C'(\cdot, i) : \mathcal{U} \to \mathbb{R}_{++}$ is a strictly increasing homeomorphism (by Assumption DARA), the Continuous Mapping Theorem (for convergence in probability) applied to the inverse of $C'(\cdot, i)$ delivers that $u_i^{(t)} \to -\infty$ in probability. *Q.E.D.*

Wrapping Up. We now complete the proof of Theorem 1.

PROOF OF THEOREM 1: We use Lemma C.23 to prove each part of the theorem in turn.

Part (a). Let $i \in S$ be given. Define $\mathbf{w}_i^{(t)} := \xi^c(\mathbf{v}^{(t)}, s^{(t)}, i)$. The promise keeping constraint (PK$_i$) requires that $v_i^{(t)} = u_i^{(t)} + \alpha\, \mathbf{E}^{\mathbf{f}_i}[\mathbf{w}_i^{(t)}]$. Because $\mathbf{w}_i^{(t)} \in D \subseteq \mathcal{U}^d$ (by definition of the domain) and $\mathcal{U} = (-\infty, 0)$ (by Assumption DARA), it follows that $v_i^{(t)} \leq u_i^{(t)}$. Lemma C.23 then implies that $v_i^{(t)} \to -\infty$ in probability, as desired.

Part (b). Immediate from Lemma C.23.

Part (c). As $u^{(t)} = U(c^{(t)} + \omega^{(t)})$ by definition and $U : (\underline{c}, \infty) \to \mathcal{U}$ is a strictly increasing homeomorphism (by Assumption DARA), the result follows from part (b) and the Continuous Mapping Theorem (for convergence in probability) applied to $U^{-1}(\cdot)$.

Nonexistence of limiting and stationary distributions. The nonexistence of limiting distributions follows directly from parts (a)–(c), proved above. For instance, for any compact $K \subseteq D$, part (a) implies that $\lim_{t\to\infty} \mathbf{P}(\mathbf{v}^{(t)} \in K) = 0$, from which it follows that $\mathbb{P}(\cdot) := \lim_{t\to\infty} \mathbf{P}(\mathbf{v}^{(t)} \in \cdot)$ cannot define a (Borel) probability measure on D.[14] The arguments for the $u^{(t)}$ and $c^{(t)} + \omega^{(t)}$ processes are analogous. We show by contradiction

[14]Note that D is σ-compact, i.e., can be written as the countable union of compact subsets $K \subseteq D$.

that there does not exist a stationary distribution for the (time-homogeneous, Markovian) state process $(\mathbf{v}^{(t)}, s^{(t)})$. Let $\Psi : D \times S \to \Delta(D \times S)$ denote the Markov kernel induced by P and the optimal contract, i.e., for all $t \in \mathbb{N}$, $(\mathbf{v}, s) \in D \times S$, and Borel $A \subseteq D \times S$, $\Psi((\mathbf{v}, s), A) := \mathsf{P}\left((\mathbf{v}^{(t+1)}, s^{(t+1)}) \in A \mid (\mathbf{v}^{(t)}, s^{(t)}) = (\mathbf{v}, s)\right)$. By induction, for every $k, t \in \mathbb{N}$, the k-step transition probabilities are given by

$$\begin{aligned}\Psi^k((\mathbf{v}, s), A) &:= \int_{D\times S} \Psi^{k-1}((\mathbf{v}', s'), A)\Psi((\mathbf{v}, s), \mathrm{d}(\mathbf{v}', s')) \\ &= \mathsf{P}\left((\mathbf{v}^{(t+k)}, s^{(t+k)}) \in A \mid (\mathbf{v}^{(t)}, s^{(t)}) = (\mathbf{v}, s)\right).\end{aligned}$$

Suppose there exists a stationary distribution $\nu \in \Delta(D \times S)$, i.e., for all Borel $A \subseteq D \times S$, $\nu(A) = \int_{D\times S} \Psi((\mathbf{v}, s), A)\, \mathrm{d}\nu(\mathbf{v}, s)$. Induction and Fubini's Theorem imply that, for every $k \in \mathbb{N}$, $\nu(A) = \int_{D\times S} \Psi^k((\mathbf{v}, s), A)\, \mathrm{d}\nu(\mathbf{v}, s)$. For every compact $A \subseteq D \times S$, part (a) and Dominated Convergence imply that $\nu(A) = \int_{D\times S} \lim_{k\to\infty} \Psi^k((\mathbf{v}, s), A)\, \mathrm{d}\nu(\mathbf{v}, s) = 0$. But this contradicts $\nu \in \Delta(D \times S)$, as desired. *Q.E.D.*

APPENDIX D: PROOF OF COROLLARY 4.1

Herein, we assume that the environment is (TVC)-Regular (as in the statement of Theorem 1) and continue to use the notation developed in SA–C (especially SA–C.4) above. To prove Corollary 4.1, it suffices to show that the vector of multipliers $\boldsymbol{\nu}^{(t)} \to \mathbf{0}$ almost surely; once we have shown this, a simple adaptation of the arguments from the proof of Lemma C.23 and the final step in the proof of Theorem 1 (with "almost surely" replacing "in probability" everywhere the latter appears) implies that promised utility, flow utility, and consumption all converge almost surely. This adaptation is straightforward, so we omit the details.

Thus, we focus on showing that $\boldsymbol{\nu}^{(t)} \to \mathbf{0}$ almost surely. To this end, note that plugging the expression for the directional derivative $\mathsf{D}_\mathbf{1} P(\mathbf{w}_i(\mathbf{v}, s), i)$ in (C.2) (from SA-C.2) into (FOC$\mathbf{w}_{ij}$) implies that, for every $(\mathbf{v}, s) \in D \times S$ and $i \in S$, we have:

$$f_{si}\left(\frac{P_j(\mathbf{w}_i(\mathbf{v}, s), i)}{f_{ij}} - \mathsf{D}_\mathbf{1} P(\mathbf{w}_i(\mathbf{v}, s), i)\right) = \sum_{k=i+1}^{d}\left(1 - \frac{f_{kj}}{f_{ij}}\right)\mu_{ki}(\mathbf{v}, s). \tag{D.1}$$

We now use (D.1) to prove each part of the corollary in turn.

PROOF OF PART (I): By hypothesis, there exists $\boldsymbol{\pi} \in \Delta(S)$ such that $f_{ij} = \pi_j$ for all $i, j \in S$.[15] Plugging this into (D.1) and invoking full connectedness (Assumption Markov) yields

$$\frac{P_j(\mathbf{w}_i(\mathbf{v}, s), i)}{f_{ij}} = \mathsf{D}_\mathbf{1} P(\mathbf{w}_i(\mathbf{v}, s), i) \tag{D.2}$$

for all $(\mathbf{v}, s) \in D \times S$ and $i, j \in S$.[16] By construction, $\mathbf{v}^{(t+1)} = \mathbf{w}_{s^{(t+1)}}(\mathbf{v}^{(t)}, s^{(t)})$ for all $t \in \mathbb{N}$. Combined with (D.2) and full connectedness (Assumption Markov), this implies that the martingale $\mathsf{D}_\mathbf{1} P(\mathbf{v}^{(t)}, s^{(t)}) \to 0$ almost surely if and only if the derivative $\mathsf{D} P(\mathbf{v}^{(t)}, s^{(t)}) \to \mathbf{0}$

[15]This notation is consistent with that from the proof of Lemma C.22: for i.i.d. Markov chains, the transition probability $\boldsymbol{\pi} \in \Delta(S)$ equals the chain's stationary distribution.

[16]Consistent with Lemma C.7, (D.2) states that $\mathsf{D} P(\mathbf{w}_i(\mathbf{v}, s), i) \in \tilde{E}_i$ as defined in ($\tilde{\mathrm{E}}_i$). That is, in the i.i.d. case, the optimal contract solves the efficiency problem (Eff$_i$) at each step.

almost surely. Thus, Lemma C.19 implies that the derivative $\mathsf{D}P(\mathbf{v}^{(t)}, s^{(t)}) \to \mathbf{0}$ almost surely. Replicating the proof of Lemma C.20 (with τ^t and $(\mathbf{v}^{\tau^t}, s^{\tau^t} = d)$ replaced everywhere by t and $(\mathbf{v}^t, s^t)$, respectively) then establishes that $\boldsymbol{\nu}^{(t)} \to \mathbf{0}$ almost surely. *Q.E.D.*

For part (ii), wherein $d = 2$, we consider the complementary cases of FOSD types ($f_{11} \geq f_{21}, f_{22} \geq f_{12}$) and non-FOSD types ($f_{11} < f_{21}, f_{22} < f_{12}$) separately.[17]

PROOF OF PART (II), FOSD CASE: We begin with two preliminary facts ((D.3) and (D.4) below). First, because $d = 2$, (D.1) for $i = 1$ reduces to

$$f_{s1}\left(\frac{P_j(\mathbf{w}_1(\mathbf{v}, s), 1)}{f_{1j}} - \mathsf{D}_1 P(\mathbf{w}_1(\mathbf{v}, s), 1)\right) = \left(1 - \frac{f_{2j}}{f_{1j}}\right)\mu_{21}(\mathbf{v}, s).$$

Because $f_{11} \geq f_{21}, f_{22} \geq f_{12}$ and $\mu_{21}(\mathbf{v}, s) \geq 0$, the above display and full connectedness (Assumption Markov) imply that

$$\frac{P_1(\mathbf{w}_1(\mathbf{v}, s), 1)}{f_{11}} \geq \mathsf{D}_1 P(\mathbf{w}_1(\mathbf{v}, s), 1) \geq \frac{P_2(\mathbf{w}_1(\mathbf{v}, s), 1)}{f_{12}}. \tag{D.3}$$

Second, because $d = 2$, the first-order condition (FOC$\mathbf{w}_{ij}$) for $i = j = 1$ reduces to

$$f_{s1} P_1(\mathbf{w}_1(\mathbf{v}, s), 1) = f_{11}\lambda_1(\mathbf{v}, s) - f_{21}\mu_{21}(\mathbf{v}, s).$$

Because $\lambda_1(\mathbf{v}, s) = P_1(\mathbf{v}, s)$ by (Env$_i$) and $\mu_{21}(\mathbf{v}, s) \geq 0$, the above display implies that

$$P_1(\mathbf{v}, s) \geq \frac{f_{s1}}{f_{11}} \cdot P_1(\mathbf{w}_1(\mathbf{v}, s), 1). \tag{D.4}$$

We now turn to the main proof. As in the proof of Lemma C.19, fix a path $h = (s^{t+1})_{t=0}^{\infty} \in \mathcal{J}^*$ and let $\tau^t := \tau^{(t)}(h)$ and $\mathbf{v}^t := \mathbf{v}^{(t)}(h)$. Denote by $\{t_k\}_{k\in\mathbb{N}}$ the range of the sequence $(\tau^{t+1})_{t=0}^{\infty}$, enumerated so that $t_k < t_{k+1}$ for all $k \in \mathbb{N}$. The proof of Lemma C.19 delivers that $\lim_{k\to\infty} \mathsf{D}_1 P(\mathbf{v}^{t_k}, s^{t_k} = 2) = 0$. Since $\mathbf{v}^{t_k} = \mathbf{w}_2(\mathbf{v}^{t_k - 1}, s^{t_k - 1})$ for all k, Lemma C.8 and ($\tilde{\mathrm{E}}_i$) then imply that the derivative satisfies $\lim_{k\to\infty} \mathsf{D}P(\mathbf{v}^{t_k}, s^{t_k} = 2) = \mathbf{0}$.

We claim that, in fact, the derivative converges along the full sequence, namely, it holds that $\lim_{t\to\infty} \mathsf{D}P(\mathbf{v}^t, s^t) = \mathbf{0}$. To this end, for every $k \in \mathbb{N}$, define $g_k := t_{k+1} - t_k$ (which is finite by definition of $\mathcal{J}^*$). By construction: (a) for every $t \in \mathbb{N}\backslash\{t_k\}_{k\in\mathbb{N}}$, there exists some $k \in \mathbb{N}$ and $1 \leq m < g_k$ such that $t = t_k + m$, and (b) for every k, we have $s^{t_k} = 2$ and $s^{t_k + m} = 1$ for all $1 \leq m < g_k$. With these facts in hand, we establish the claim in three steps.

Step 1: We first use (D.4) to show that, for every $k \in \mathbb{N}$, it holds that

$$P_1(\mathbf{v}^{t_k}, s^{t_k} = 2) \geq \frac{f_{21}}{f_{11}} \cdot \max_{1 \leq m < g_k} P_1(\mathbf{v}^{t_k + m}, s^{t_k + m} = 1). \tag{D.5}$$

To this end, consider any date t_k such that $g_k \geq 2$. (If $g_k = 1$, there is nothing to prove.) By definition of t_k and g_k, we have $\mathbf{v}^{t_k + m} = \mathbf{w}_1(\mathbf{v}^{t_k + m - 1}, s^{t_k + m - 1})$ for all $1 \leq m < g_k$. Thus, for $m = 1$, (D.4) yields $P_1(\mathbf{v}^{t_k}, s^{t_k} = 2) \geq \frac{f_{21}}{f_{11}} \cdot P_1(\mathbf{v}^{t_k + 1}, s^{t_k + 1} = 1)$. If $g_k = 2$, this immediately yields (D.5). If $g_k > 2$, then for every $2 \leq m < g_k$, (D.4) yields $P_1(\mathbf{v}^{t_k + m - 1}, s^{t_k + m - 1} = 1) \geq P_1(\mathbf{v}^{t_k + m}, s^{t_k + m} = 1)$. Stringing these inequalities together yields (D.5).

[17]When $d = 2$, these two cases are exhaustive because the identity $f_{11} + f_{12} = 1 = f_{21} + f_{22}$ implies that $f_{11} - f_{21} = f_{22} - f_{12}$.

Step 2: Next, we use (D.3) and (D.5) to show that, for every $k \in \mathbb{N}$, it holds that

$$P_2(\mathbf{v}^{t_k}, s^{t_k} = 2) \geq \frac{f_{22}}{f_{12}} \cdot \max_{1 \leq m < g_k} P_2(\mathbf{v}^{t_k+m}, s^{t_k+m} = 1). \tag{D.6}$$

To this end, note that because $\mathbf{v}^{t_k} = \mathbf{w}_2(\mathbf{v}^{t_k-1}, s^{t_k-1})$ by definition of t_k, Lemma C.8 and ($\tilde{\mathrm{E}}_i$) imply that $P_2(\mathbf{v}^{t_k}, s^{t_k} = 2) = \frac{f_{22}}{f_{21}} P_1(\mathbf{v}^{t_k}, s^{t_k} = 2)$. Plugging this in to (D.5) yields

$$\begin{aligned} P_2(\mathbf{v}^{t_k}, s^{t_k} = 2) &\geq \frac{f_{22}}{f_{21}} \cdot \frac{f_{21}}{f_{11}} \cdot \max_{1 \leq m < g_k} P_1(\mathbf{v}^{t_k+m}, s^{t_k+m} = 1) \\ &\geq \frac{f_{22}}{f_{21}} \cdot \frac{f_{21}}{f_{11}} \cdot \frac{f_{11}}{f_{12}} \max_{1 \leq m < g_k} P_2(\mathbf{v}^{t_k+m}, s^{t_k+m} = 1), \end{aligned}$$

where the second inequality follows from (D.3) applied to each term on the RHS of the first line. Canceling terms yields (D.6).

Step 3: We now prove the claim, i.e., $\lim_{t\to\infty} \mathsf{D}P(\mathbf{v}^t, s^t) = \mathbf{0}$. Recall that (as established above) the derivative satisfies $\lim_{k\to\infty} \mathsf{D}P(\mathbf{v}^{t_k}, s^{t_k} = 2) = \mathbf{0}$. When combined with (D.5) and (D.6), this implies that the partial derivatives satisfy $\limsup_{t\to\infty} P_i(\mathbf{v}^t, s^t) \leq 0$ for $i = 1, 2$. But since the directional derivative $\mathsf{D}_1 P(\mathbf{v}^t, s^t) \geq 0$ for all t by Lemma C.1, it follows that $\lim_{t\to\infty} P_i(\mathbf{v}^t, s^t) = 0$ for all $i \in S$. That is, $\lim_{t\to\infty} \mathsf{D}P(\mathbf{v}^t, s^t) = \mathbf{0}$, as claimed.

We now complete the proof. Since $\lim_{t\to\infty} \mathsf{D}P(\mathbf{v}^t, s^t) = \mathbf{0}$, replicating the proof of Lemma C.20 (with τ^t and $(\mathbf{v}^{\tau^t}, s^{\tau^t} = d)$ replaced everywhere by t and $(\mathbf{v}^t, s^t)$, respectively) delivers $\boldsymbol{\nu}(\mathbf{v}^t, s^t) \to \mathbf{0}$. Since $\mathsf{P}(\mathcal{J}^*) = 1$, we have $\boldsymbol{\nu}^{(t)} \to \mathbf{0}$ almost surely, as desired. *Q.E.D.*

PROOF OF PART (II), NON-FOSD CASE: The argument is analogous to the FOSD case above. We begin by deriving analogues of (D.3) and (D.4) ((D.7) and (D.8) below). First, because we now have $f_{11} < f_{21}$, $f_{22} < f_{12}$, the inequalities in (D.3) flip, delivering

$$\frac{P_1(\mathbf{w}_1(\mathbf{v}, s), 1)}{f_{11}} \leq \mathsf{D}_1 P(\mathbf{w}_1(\mathbf{v}, s), 1) \leq \frac{P_2(\mathbf{w}_1(\mathbf{v}, s), 1)}{f_{12}}. \tag{D.7}$$

Second, because $d = 2$, the FOC (FOC$\mathbf{w}_{ij}$) for $i = 1$ and $j = 2$ reduces to

$$f_{s1} P_2(\mathbf{w}_1(\mathbf{v}, s), 1) = f_{12}\lambda_1(\mathbf{v}, s) - f_{22}\mu_{21}(\mathbf{v}, s).$$

Because $\lambda_1(\mathbf{v}, s) = P_1(\mathbf{v}, s)$ by (Env$_i$) and $\mu_{21}(\mathbf{v}, s) \geq 0$, this FOC and (D.7) together imply

$$P_1(\mathbf{v}, s) \geq \frac{f_{s1}}{f_{12}} P_2(\mathbf{w}_1(\mathbf{v}, s), 1) \geq \frac{f_{s1}}{f_{11}} P_1(\mathbf{w}_1(\mathbf{v}, s), 1). \tag{D.8}$$

We now turn to the main proof. Let h, τ^t, $\mathbf{v}^t$, t_k, and g_k be as defined above in the proof for the FOSD case. We obtain (D.5) using the same argument as in Step 1 from the FOSD case, except with (D.8) replacing (D.4) everywhere the latter appears. To obtain (D.6), we modify Step 2 from the FOSD case as follows:

Step 2′: For every $k \in \mathbb{N}$, Lemma C.8 and ($\tilde{\mathrm{E}}_i$) imply that $P_2(\mathbf{v}^{t_k}, s^{t_k} = 2) = \frac{f_{22}}{f_{21}} P_1(\mathbf{v}^{t_k}, s^{t_k} = 2)$. For every k such that $g_k \geq 2$ (there is nothing to prove if $g_k = 1$), plugging this into (the first inequality in) (D.8) then delivers

$$P_2(\mathbf{v}^{t_k}, s^{t_k} = 2) \geq \frac{f_{22}}{f_{12}} P_2(\mathbf{v}^{t_k+1}, s^{t_k+1} = 1).$$

If $g_k = 2$, this immediately yields (D.6). If $g_k > 2$, then for every $2 \le m < g_k$, we have

$$\begin{aligned} P_2(\mathbf{v}^{t_k+m-1}, s^{t_k+m-1} = 1) &\ge \frac{f_{12}}{f_{11}} P_1(\mathbf{v}^{t_k+m-1}, s^{t_k+m-1} = 1) \\ &\ge \frac{f_{12}}{f_{11}} \cdot \frac{f_{11}}{f_{12}} P_2(\mathbf{v}^{t_k+m}, s^{t_k+m} = 1) = P_2(\mathbf{v}^{t_k+m}, s^{t_k+m} = 1), \end{aligned}$$

where the first inequality is by (D.7), the second inequality is by (the first inequality in) (D.8), and the final equality is by canceling terms. Stringing the inequalities in the above two displays together yields (D.6).

Step 3 and the final paragraph from the proof of the FOSD case then carry over verbatim, completing the present proof. *Q.E.D.*

APPENDIX E: PROOF OF THEOREM 2

We first prove Theorem 3(d) (from Appendix B.1), which we then use to prove Theorem 2.

E.1. *Proof of Theorem 3(d)*

For every $i, k \in S$, denote by $F^*(k \mid i) := \sum_{\ell \ge k} f_{i\ell}$ the probability of transitioning from type i to some type weakly greater than k. For every $i \in S \backslash \{1\}$, define $\Delta F^*_{i,i-1}(k) := F^*(k \mid i) - F^*(k \mid i-1)$. By FOSD, $\Delta F^*_{i,i-1}(k) \ge 0$ for all $i \in S\backslash\{1\}$ and $k \in S$ (with equality for $k = 1$). For every $i \in S\backslash\{1\}$ and $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, a standard calculation delivers

$$\mathsf{E}^{\mathbf{f}_i}[\mathbf{x}] - \mathsf{E}^{\mathbf{f}_{i-1}}[\mathbf{x}] = \sum_{k=1}^{d} (f_{ik} - f_{i-1,k}) x_k = \sum_{k=1}^{d-1} \Delta F^*_{i,i-1}(k+1)[x_{k+1} - x_k]. \tag{E.1}$$

Now, recall the sets $D^* \subseteq D$ and $\Xi^*(\mathbf{v}) \subseteq \Xi$ defined in Appendix B.1. Let $\mathbf{v} \in D^*$ and a contract $\xi \in \Xi^*(\mathbf{v})$ be given. Let $(\mathbf{v}^{(t)})_{t=0}^{\infty}$ be the induced promise processes (initialized at $\mathbf{v}^{(0)} := \mathbf{v}$) and let $(u_i^{(t)}, \mathbf{w}_i^{(t)}) := \xi(\mathbf{v}^{(t)}, s^{(t)}, i)$ for each $i \in S$.[18] For every $t \in \mathbb{N} \cup \{0\}$ and $i \in S\backslash\{1\}$, the (IC^*_{ij}) constraint (for $j = i-1$) yields

$$\begin{aligned} v_i^{(t)} - v_{i-1}^{(t)} &\ge \psi(u_{i-1}^{(t)}, i, i-1) - u_{i-1}^{(t)} + \alpha \Big[\mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_{i-1}^{(t)}] - \mathsf{E}^{\mathbf{f}_{i-1}}[\mathbf{w}_{i-1}^{(t)}] \Big] \\ &= Z_i(u_{i-1}^{(t)}) + \alpha \sum_{k=1}^{d-1} \Delta F^*_{i,i-1}(k+1)[w^{(t)}_{i-1,k+1} - w^{(t)}_{i-1,k}], \end{aligned}$$

where in the second line we let $Z_i(u) := \psi(u, i, i-1) - u$ for all $u \in \mathcal{U}$ and invoke (E.1) for $\mathbf{x} = \mathbf{w}_{i-1}^{(t)}$. For every $i \in S\backslash\{1\}$ and $t \in \mathbb{N}$, let $\hat{H}_i^t := \{h^t = (s^1, \dots, s^t) \in S^t : s^1 = i-1,\ s^\tau \ne d\ \forall \tau = 1, \dots, t\}$. Then, for each $i \in S\backslash\{1\}$, iterating the above display forward T times starting from $t = 0$ yields

$$v_i - v_{i-1} \ge Z_i(u_{i-1}^{(0)}) + \sum_{t=1}^{T} \alpha^t \sum_{h^{t+1} \in \hat{H}_i^{t+1}} \left(\prod_{\tau=1}^{T} \Delta F^*_{s^\tau+1, s^\tau}(s^{\tau+1}+1) \right) Z_{s^{t+1}+1}(u^{(t)}_{s^{t+1}}(h^t))$$

[18] Two remarks on notation: (i) we drop the ξ subscript on the induced promises for simplicity, and (ii) unlike in SA–C above, ξ need not be optimal and the policies $(u_i^{(t)}, \mathbf{w}_i^{(t)})$ need not satisfy the FOCs.

$$+\alpha^{T+1}\sum_{h^{T+2}\in\hat{H}_i^{T+2}}\left(\prod_{\tau=1}^{T+1}\Delta F^*_{s^\tau+1,s^\tau}(s^{\tau+1}+1)\right)\left[v^{(T+1)}_{s^{T+2}+1}(h^{T+1})-v^{(T+1)}_{s^{T+2}}(h^{T+1})\right],$$

where at each step of the iteration we use the identity $\mathbf{v}^{(t+1)}((h^t,j))=\mathbf{w}^{(t)}_j(h^t)$ for all $h^t\in S^t$ and $j\in S$. Note that the first line of the RHS is bounded below by some $M_i>0$ because $Z_k(\cdot)>0$ for all $k\in S\backslash\{1\}$, and FOSD implies that each $\Delta F^*_{s^\tau+1,s^\tau}(s^{\tau+1}+1)\geq 0$. To bound below the second line of the RHS, let $\varepsilon>0$ be given and note that, since $\xi\in\Xi^*(\mathbf{v})$ and hence $\lim_{t\to\infty}\inf_{h\in\mathcal{H}}\alpha^t\mathbf{v}^{(t)}(h)=\mathbf{0}$ (i.e., (TVC) holds), there exists $T_\varepsilon\in\mathbb{N}$ such that, for all $T\geq T_\varepsilon$ and paths $h\in H$, we have $\alpha^T\mathbf{v}^{(T)}(h)\geq-\varepsilon\mathbf{1}$ and hence $\alpha^T\left(v^{(T)}_{i+1}(h)-v^{(T)}_i(h)\right)\geq-\varepsilon$ for all $i\in S\backslash\{d\}$. Plugging these bounds into the above display and using FOSD on the second line, we conclude that, for all $T\geq T_\varepsilon$,

$$v_i-v_{i-1}\geq M_i-\varepsilon\cdot\sum_{h^{T+2}\in\hat{H}_i^{T+2}}\left(\prod_{\tau=1}^{T+1}\Delta F^*_{s^\tau+1,s^\tau}(s^{\tau+1}+1)\right).\tag{E.2}$$

By repeated use of (E.1) and the Law of Iterated Expectations, backward induction yields

$$\sum_{h^{T+2}\in\hat{H}_i^{T+2}}\left(\prod_{\tau=1}^{T+1}\Delta F^*_{s^\tau+1,s^\tau}(s^{\tau+1}+1)\right)=\mathsf{E}[s^{(T+2)}\mid s^{(1)}=i]-\mathsf{E}[s^{(T+2)}\mid s^{(1)}=i-1],$$

which vanishes as $T\to\infty$ because the type process is ergodic. Thus, sending $T\to\infty$ in (E.2) delivers $v_i-v_{i-1}\geq M_i>0$. We conclude that $\mathbf{v}\in V_d$. Hence, $D^*\subseteq V_d$, as desired.

E.2. *Main Proof of Theorem 2*

Herein, we maintain the hypotheses that the environment is (TVC)-Regular and the type process is FOSD, and continue to use the notation developed in SA–C above.

PROOF OF PART (A): Let $i\in S\backslash\{1\}$ be given. Since the environment is (TVC)-Regular, Proposition 3.2(b) implies that the unique optimal contract is (TVC)-implementable. Thus, the continuation utility process $\mathbf{w}^{(t)}_{i-1}:=\xi^c(\mathbf{v}^{(t)},s^{(t)},i-1)$ satisfies $\mathbf{w}^{(t)}_{i-1}\in D^*$ pathwise. Since the type process is FOSD, Theorem 3(d) then implies that $\mathbf{w}^{(t)}_{i-1}\in V_d$, i.e., the map $k\mapsto w^{(t)}_{i-1,k}$ is increasing. Thus, FOSD implies that $\mathsf{E}^{\mathbf{f}_i}\left[\mathbf{w}^{(t)}_{i-1}\right]-\mathsf{E}^{\mathbf{f}_{i-1}}\left[\mathbf{w}^{(t)}_{i-1}\right]\geq 0$ pathwise. Plugging this inequality into the IC constraint (IC^*_{ij}) (for $j=i-1$) from Section 4.2 yields

$$\begin{aligned}v^{(t)}_i-v^{(t)}_{i-1}&\geq\psi(u^{(t)}_{i-1},i,i-1)-u^{(t)}_{i-1}+\alpha\left(\mathsf{E}^{\mathbf{f}_i}\left[\mathbf{w}^{(t)}_{i-1}\right]-\mathsf{E}^{\mathbf{f}_{i-1}}\left[\mathbf{w}^{(t)}_{i-1}\right]\right)\\&\geq\psi(u^{(t)}_{i-1},i,i-1)-u^{(t)}_{i-1}.\end{aligned}\tag{E.3}$$

Define $Z_i(u):=\psi(u,i,i-1)-u$. Assumption DARA implies that the map $Z_i:\mathcal{U}\to(0,\infty)$ is strictly decreasing, convex, and continuously differentiable.[19] Let $a_i:=Z_i'(-1)<0$ and

[19]The Inverse Function Theorem delivers $Z_i'(u)=\frac{U'(\omega_i-\omega_{i-1}+U^{-1}(u))}{U'(U^{-1}(u))}-1$. Since $\omega_i>\omega_{i-1}$ and U is strictly concave, $Z_i'(u)<0$. Part (c) of Assumption DARA implies that the map $u\mapsto-\log\left(Z_i'(u)+1\right)$ is weakly decreasing. Thus, $Z_i'(\cdot)$ is non-decreasing, i.e., $Z_i(\cdot)$ is convex.

$b_i := Z_i(-1) + a_i \in \mathbb{R}$. By convexity, $Z_i(u) \geq Z_i(-1) + (u+1) \cdot Z_i'(-1) = a_i u + b_i$ for all $u \in \mathcal{U}$. Plugging this inequality into (E.3) delivers

$$v_i^{(t)} - v_{i-1}^{(t)} \geq Z_i(u_{i-1}^{(t)}) \geq a_i \cdot u_i^{(t)} + b_i \to +\infty \text{ in probability,}$$

where the limit is by $a_i < 0$ and Lemma C.23. This completes the proof. *Q.E.D.*

PROOF OF PART (B): Let $\overline{f}_d := \max_{s \in S} f_{sd}$ and $\underline{f}_d := \min_{s \in S} f_{sd}$. By the same argument as in the proof of part (a) above, the promised utility process satisfies $\mathbf{v}^{(t)} \in D^* \subseteq V_d$, and thus

$$v_d^{(t)} \geq \overline{f}_d v_d^{(t)} + (1 - \overline{f}_d) v_{d-1}^{(t)} \geq \sum_{i=1}^{d} f_{s^{(t)},i} v_i^{(t)}$$

along every path. Plugging the above into (4.1) from Section 4.2, we obtain

$$\begin{aligned} \mathsf{V}\left(v_{s^{(t+1)}}^{(t)} \mid \mathbf{v}^{(t)}, s^{(t)}\right) &= \sum_{i=1}^{d} f_{s^{(t)},i} \left(v_i^{(t)} - \sum_{i=1}^{d} f_{s^{(t)},i} v_i^{(t)}\right)^2 \\ &\geq \underline{f}_d \left(v_d^{(t)} - \sum_{i=1}^{d} f_{s^{(t)},i} v_i^{(t)}\right)^2 \geq \underline{f}_d (1 - \overline{f}_d)^2 \cdot (v_d^{(t)} - v_{d-1}^{(t)})^2. \end{aligned}$$

Full connectedness (Assumption Markov) yields $\overline{f}_d, \underline{f}_d \in (0,1)$. Thus, part (a) implies $\mathsf{V}\left(v_{s^{(t+1)}}^{(t)} \mid \mathbf{v}^{(t)}, s^{(t)}\right) \to +\infty$ in probability, as desired. *Q.E.D.*

APPENDIX F: FIRST-BEST BENCHMARK

Herein, we characterize the first-best contract that is optimal under full information. We adopt a recursive formulation analogous to that for the second-best problem in Section 3.1, the only difference being that the incentive constraints (IC_{ij}) are no longer included.[20]

A *(recursive) full-information contract* is a map $\zeta : \mathcal{U}^d \times S \to (\mathcal{U} \times \mathcal{U}^d)^d$, and we say that ζ is *feasible* if $\zeta(\mathbf{v}, s) \in \Gamma^{\mathrm{FB}}(\mathbf{v})$ for every $(\mathbf{v}, s) \in \mathcal{U}^d \times S$, where $\Gamma^{\mathrm{FB}}(\mathbf{v}) := \left\{(u_i, \mathbf{w}_i)_{i \in S} \in (\mathcal{U} \times \mathcal{U}^d)^d : (\mathrm{PK}_i) \text{ holds } \forall\, i \in S \text{ at } \mathbf{v} \in \mathcal{U}^d\right\}$.[21] Let Ξ^{FB} denote the set of feasible full-information contracts. The principal's *full-information recursive problem* is

$$Q^*(\mathbf{v}, s) := \inf_{\zeta \in \Xi^{\mathrm{FB}}} \mathsf{E}\left[\sum_{t=0}^{\infty} \alpha^t C\left(u_\zeta^{(t)}, s^{(t+1)}\right) \,\middle|\, (\mathbf{v}_\zeta^{(0)}, s^{(0)}) = (\mathbf{v}, s)\right], \tag{FB}$$

where the processes of induced allocations $(u_\zeta^{(t)})_{t=0}^\infty$ and induced promises $(\mathbf{v}_\zeta^{(t)})_{t=0}^\infty$ are defined by iterating on ζ in the natural way (cf. the recursive problem (RP) in Section 3.1). A contract ζ^* is *first-best* if it attains the infimum in (FB) at every $(\mathbf{v}, s) \in \mathcal{U}^d \times S$. We represent the infimal cost in (FB) by the principal's *first-best value function* $Q^* : \mathcal{U}^d \times S \to \overline{\mathbb{R}}$.

[20]Under Condition R.2, this recursive formulation is equivalent to the full-information analogue of the sequential formulation in Appendix A because the first-best contract in Lemma F.1 below satisfies (TVC).

[21]In the full-information problem, every $\mathbf{v} \in \mathcal{U}^d$ is implementable (e.g., via the first-best contract in Lemma F.1(a) below). Thus, the (recursive) domain for the full-information problem is the entirety of $\mathcal{U}^d$.

LEMMA F.1: *Suppose that Condition R.2 holds. Then the first-best value function $Q^* : \mathcal{U}^d \times S \to \mathbb{R}$ is finite-valued and satisfies the functional equation*

$$Q^*(\mathbf{v}, s) = \min_{(u_i, \mathbf{w}_i)_{i\in S} \in \Gamma^{FB}(\mathbf{v})} \sum_{i\in S} f_{si}\left[C(u_i, i) + \alpha Q^*(\mathbf{w}_i, i)\right]. \tag{F.1}$$

Furthermore, $Q^*(\cdot, s)$ is strictly convex, strictly increasing in the direction $\mathbf{1}$, and continuously differentiable for each $s \in S$. Moreover:

(a) *There exists a unique first-best contract, which is given by $\zeta^*(\mathbf{v}, s) = ((1-\alpha)v_i, v_i\mathbf{1})_{i\in S}$ for every $(\mathbf{v}, s) \in \mathcal{U}^d \times S$.*

(b) *The full-information contract generated by the (unique) policy function from (F.1) is ζ^*.*

PROOF: For each $s \in S$, $Q^*(\cdot, s) < +\infty$ on $\mathcal{U}^d$ because the contract described in part (a) is feasible and has finite cost, and $Q^*(\cdot, s)$ is convex by (FB) because each $C(\cdot, i)$ is convex and $\Gamma^{\text{FB}} : \mathcal{U}^d \rightrightarrows (\mathcal{U} \times \mathcal{U}^d)^d$ has convex graph. Thus, $Q^*(\mathbf{v}, s) = -\infty$ for some $(\mathbf{v}, s) \in \mathcal{U}^d \times S$ only if $Q^*(\mathbf{v}', s) = -\infty$ for all $\mathbf{v}' \in \mathcal{U}^d$, which would violate Condition R.2. Hence, each $Q^*(\cdot, s)$ is finite-valued and convex. Standard arguments then imply that Q^* satisfies (F.1), the minimum in (F.1) is attained and, under Condition R.2, the policy functions of (F.1) generate first-best contracts, which therefore exist (cf. Lemmas J.7–J.10 in Section J of Bloedel, Krishna, and Leukhina (2024)). It is easy to see from (FB) that each $Q^*(\cdot, s)$ is non-decreasing in the direction $\mathbf{1}$. Moreover, the Benveniste and Scheinkman (1979) envelope theorem applied to (F.1) yields that each $Q^*(\cdot, s) \in \mathbf{C}^1(\mathcal{U}^d)$.[22] To complete the proof, it suffices to show that the contract described in part (a) is first-best; given this fact, the strict convexity and monotonicity of each $C(\cdot, i)$ imply via (F.1) that each $Q^*(\cdot, s)$ is strictly convex and strictly increasing in the direction $\mathbf{1}$, and strict convexity of each $Q^*(\cdot, s)$ then yields the uniqueness claims in parts (a) and (b).

Thus, we claim that the contract $\zeta^*(\mathbf{v}, s) := ((1-\alpha)v_i, v_i\mathbf{1})_{i\in S}$, which is feasible by construction, is first-best. To this end, let ζ be any first-best contract and let $(\mathbf{v}^{(0)}, s^{(0)}) \in \mathcal{U}^d \times S$ be given. Standard arguments imply that the induced allocation $u_\zeta^{(t)}$ under ζ is a constant process *conditional on the initial type* $s^{(1)}$, i.e., conditional on $s^{(1)} = i$, there exists some $z_i(\mathbf{v}^{(0)}, s^{(0)}) \in \mathcal{U}$ such that $u_\zeta^{(t)} = z_i(\mathbf{v}^{(0)}, s^{(0)})$ for all $t \geq 0$.[23] Iterating the (PK$_i$) constraints forward in time implies that, for each $i \in S$, we have

$$v_i^{(0)} = \lim_{T\to\infty}\left[\frac{1-\alpha^{T+1}}{1-\alpha} z_i(\mathbf{v}^{(0)}, s^{(0)}) + \alpha^{T+1}\mathbf{E}\left[\mathbf{E}^{\mathbf{f}_{s^{(T+1)}}}[\mathbf{v}_\zeta^{(T+1)}] \mid s^{(1)} = i\right]\right] \leq \frac{z_i(\mathbf{v}^{(0)}, s^{(0)})}{1-\alpha}$$

where the inequality is by $\mathcal{U} = (-\infty, 0)$ (Assumption DARA). Then, since ζ^* is feasible and costs weakly more than ζ (which is first-best), it must be that $z_i(\mathbf{v}^{(0)}, s^{(0)}) = (1-\alpha)v_i^{(0)}$ for each $i \in S$, i.e., ζ^* and ζ induce the same allocations starting from $(\mathbf{v}^{(0)}, s^{(0)})$. Hence, ζ^* also attains the optimal value $Q^*(\mathbf{v}^{(0)}, s^{(0)})$ starting from $(\mathbf{v}^{(0)}, s^{(0)})$. Since the initial condition was arbitrary, we conclude that ζ^* is first-best, as claimed. *Q.E.D.*

[22]Formally, let $(\mathbf{v}, s) \in \mathcal{U}^d \times S$ be given. Since $Q^*(\cdot, s)$ is convex, it suffices to show that each partial derivative exists at $\mathbf{v}$. So, let $j \in S$ be given. Define $\mathbf{v}(t) := \mathbf{v} + t\hat{e}_j$ where $\hat{e}_j \in \mathbb{R}^d$ is the unit vector in the jth direction. Given a menu $(u_i, \mathbf{w}_i)_{i\in S} \in \Gamma^{\text{FB}}(\mathbf{v})$ that attains the minimum in (F.1) at $(\mathbf{v}, s)$, define $u_i(t) := u_i + t\mathbf{1}(i = j)$ and $\mathbf{w}_i(t) := \mathbf{w}_i$ for each $i \in S$. Since $\mathcal{U}$ is open, there is an $\varepsilon > 0$ such that $\mathbf{v}(t) \in \mathcal{U}^d$ and $u_j(t) \in \mathcal{U}$ for $|t| < \varepsilon$. By construction, $(u_i(t), \mathbf{w}_i(t))_{i\in S} \in \Gamma^{\text{FB}}(\mathbf{v}(t))$ for all such t. Then, since $C(\cdot, j) \in \mathbf{C}^1(\mathcal{U})$, Benveniste and Scheinkman (1979) implies that the partial derivative $Q_j^*(\mathbf{v}, s)$ exists.

[23]This can be seen either from (FB) directly or from the envelope and first-order conditions for (F.1) (which mirror those for the second-best problem in Appendix C.1, except with Q^* appearing in place of P and $\mu_{ij}(\mathbf{v}, s) := 0$ for all $i, j \in S$).

Lemma F.1(a) shows that the first-best contract is unique and fully insures the agent *conditional on his initial type*. Specifically, conditional on $\omega^{(0)} = \omega_i$ (i.e., $s^{(1)} = i$), the induced allocations $u^{(t)}_{\xi^*} \equiv (1-\alpha)v^{(0)}_i$ are constant for $t \geq 0$ and the induced promises $\mathbf{v}^{(t)}_{\xi^*} \equiv v^{(0)}_i \mathbf{1}$ are constant for $t \geq 1$. However, the first-best contract fully insures the agent against his initial type if and only if the initial $\mathbf{v}^{(0)}$ is on the diagonal (i.e., $v^{(0)}_1 = \cdots = v^{(0)}_d$).

Naturally, if the principal could choose the initial $\mathbf{v}^{(0)}$, she would choose it to lie on the diagonal. To model this choice, consider for each $i \in S$ the *first-best efficiency problem*:

$$K^*(v,i) := \min_{\mathbf{v} \in \mathcal{U}^d} Q^*(\mathbf{v},i) \quad \text{s.t.} \quad \mathsf{E}^{\mathbf{f}_i}[\mathbf{v}] \geq v. \tag{Eff$_i^{\mathrm{FB}}$}$$

This is the full-information analogue of the efficiency problem (Eff$_i$) from Appendix C.3.1.

LEMMA F.2: *Suppose that Condition R.2 holds. Then, for each $i \in S$, the unique solution $\mathbf{v}^*(v,i) \in \mathcal{U}^d$ and corresponding value of (Eff$_i^{\mathrm{FB}}$) at $v \in \mathcal{U}$ are given by*

$$\mathbf{v}^*(v,i) = v \cdot \mathbf{1} \quad \text{and} \quad K^*(v,i) = \frac{U^{-1}((1-\alpha)v)}{1-\alpha} - \mathsf{E}\left[\sum_{t=0}^{\infty} \alpha^t \omega^{(t)} \,\middle|\, s^{(0)} = i\right]. \tag{F.2}$$

Moreover, the value function $K^*(\cdot,i) : \mathcal{U} \to \mathbb{R}$ is strictly increasing, strictly convex, continuously differentiable, unbounded above (i.e., $\lim_{v\to 0} K^*(v,i) = +\infty$), and satisfies the Inada conditions $\lim_{v\to-\infty}(K^*)'(v,i) = 0$ and $\lim_{v\to 0}(K^*)'(v,i) = +\infty$.

PROOF: Existence of a solution to (Eff$_i^{\mathrm{FB}}$) follows from routine arguments (cf. Lemma J.10 in Section J of Bloedel, Krishna, and Leukhina (2024)). Uniqueness of the solution, the expressions in (F.2), and the strict monotonicity, strict convexity, and continuous differentiability of $K^*(\cdot,i)$ then follow from Lemma F.1. The limiting properties of $K^*(\cdot,i)$ follow from (F.2) and the fact that $v \mapsto U^{-1}((1-\alpha)v)$ satisfies the same properties (by Assumption DARA). *Q.E.D.*

APPENDIX G: PATHWISE PROPERTIES OF MARKOV CHAINS

Herein, we collect facts about the paths of Markov chains, which we use in the proof of Theorem 1 (see SA-C.4). The first fact is standard; see Shiryaev (1995, p. 577) for a proof.

LEMMA G.1: *Let $(X^{(t)})$ be a time-homogeneous Markov chain with countable state space $\mathcal{X}$ and law $\mathbb{P} \in \Delta(\mathcal{X}^\infty)$ over paths. If state $x \in \mathcal{X}$ is recurrent, then $\mathbb{P}(X^{(t)} = x$ for infinitely many $t \mid X^{(0)} = x) = 1$.*

The next result applies Lemma G.1 to our setting.

LEMMA G.2: *For all $i, j \in S$, we have $\mathsf{P}\big((s^{(t-1)}, s^{(t)}) = (i,j)$ for infinitely many $t\big) = 1$.*

PROOF: Consider the time-homogeneous Markov chain $X^{(t)} := (s^{(t)}, s^{(t+1)})$ with finite state space $\mathcal{X} := S \times S$, and law $\mathbb{P}$ induced by the initial distribution of $s^{(0)}$ under P and the transition probabilities $Q : \mathcal{X} \to \Delta(\mathcal{X})$ generated by P via $Q((i,j),(k,\ell)) := \mathbf{1}(j = k) \cdot f_{k\ell}$. The two-step transition probabilities are then $Q^{(2)}\big((i,j),(k,\ell)\big) = f_{jk} f_{k\ell} > 0$, so the chain is indecomposable. Let $i, j \in S$ be given. It follows that $\tau(i,j) := \inf\{t \in \mathbb{N} : X^{(t)} = (i,j)\}$ is finite $\mathbb{P}$-a.s. Then, Lemma G.1 and the Strong Markov Property (e.g., Norris (1997, Theorem 1.2, p. 20)) imply that $\mathbb{P}(X^{(k+\tau(i,j))} = (i,j)$ for infinitely many $k \mid X^{(\tau(i,j))} = (i,j)) = 1$. It follows that $\mathbb{P}\left(X^{(t)} = (i,j) \text{ for infinitely many } t\right) = 1$, as desired. *Q.E.D.*

The final result is a direct corollary of Lemma G.2 and the fact that the intersection of finitely-many full-measure events has full measure.

COROLLARY G.3: *Define the events* $\mathcal{I}_j, \mathcal{I} \subseteq \mathcal{H}$ *as*

$$\mathcal{I}_j := \{h \in \mathcal{H} : (s^{t-1}, s^t) = (d, j) \text{ infinitely often}\} \quad \text{for all } j \in S$$

and $\mathcal{I} := \bigcap_{j=1}^{d} \mathcal{I}_j$. *Then,* $\mathbf{P}(\mathcal{I}_j) = 1$ *for all* $j \in S$ *and hence* $\mathbf{P}(\mathcal{I}) = 1$.

APPENDIX H: DISCUSSION OF CONDITION R.5

Herein, we expand on the discussion of Condition R.5 in Section 3.2. We assume that Conditions R.1–R.4 hold (but do *not* assume R.5). Thus, P satisfies the Bellman equation (FE) and is strictly convex, and there exists a unique optimal contract (Proposition 3.2).[24]

Because our constraint set $\Gamma(\mathbf{v})$ includes both IC constraints and multiple interim promise keeping constraints, the standard envelope theorem for concave dynamic programs of Benveniste and Scheinkman (1979) does not directly apply.[25] We describe an alternative envelope theorem of Rincón-Zapatero and Santos (2009, Theorem 3.1) (henceforth RZS), which permits constraints of this form.

RZS's setup applies to our "interim" Bellman equation (FE-Q^i) from SA–C.3.3, once we have solved (PK$_i$) for the flow utility $u_i = v_i - \alpha\,\mathbf{E}^{\mathbf{f}_i}[\mathbf{w}_i]$, so that $\mathbf{w}_i \in D$ is the only choice variable. Substituting this into (IC$^*_{ji}$) yields the reduced-form incentive constraints

$$g^j(\mathbf{v}, \mathbf{w}_i, i) := v_j - \alpha\,\mathbf{E}^{\mathbf{f}_j}[\mathbf{w}_i] - \psi\left(v_i - \alpha\,\mathbf{E}^{\mathbf{f}_i}[\mathbf{w}_i], j, i\right) \geq 0 \qquad (\text{RIC}^*_{ji})$$

for all $j > i \in S$. We denote the derivative of $g^j(\mathbf{v}, \cdot, i)$ by

$$\nabla_2 g^j(\mathbf{v}, \mathbf{w}_i, i) = \alpha\left\{-\mathbf{f}_j + \psi'\left(v_i - \alpha\,\mathbf{E}^{\mathbf{f}_i}[\mathbf{w}_i], j, i\right)\mathbf{f}_i\right\} \in \mathbb{R}^d. \qquad (\text{H.1})$$

RZS require four technical conditions, D1–D4, to conclude that Condition R.5 holds.

- D1 requires that the principal's flow cost $C(\cdot, i)$ is $\mathbf{C}^1$. This holds in our setting because the agent's utility function U is $\mathbf{C}^1$ (Assumption DARA).
- D2 requires that, under the optimal contract, $(\mathbf{v}^{(t)})_{t=0}^{\infty}$ evolves in the interior of the domain D. This holds in our setting because D is open (Theorem 3).
- D3 is a constraint qualification requiring that for every $\mathbf{v} \in D$ and $i \in S$, *at the optimal* $\mathbf{w}_i(\mathbf{v})$ the derivatives corresponding to binding constraints $\{\nabla_2 g^j(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) : g^j(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) = 0\}_{j=i+1}^{d}$ are linearly independent, i.e., the matrix formed by using these

[24]In Proposition 3.2, Condition R.5 is only used to establish that $P(\cdot, s) \in \mathbf{C}^1(D)$ in part (b).

[25]With a single *ex ante* promised utility state $v \in \mathcal{U}$ and corresponding single *ex ante* promise keeping constraint, as in standard recursive formulations of the i.i.d. case, the Benveniste and Scheinkman (1979) theorem typically applies because we can implement any perturbation of v by varying only the flow utilities $(u_i)_{i \in S}$ and the flow cost $C(\cdot, i)$ is smooth. See Footnote 34 below for details. By contrast, with a vector of interim promised utilities $\mathbf{v} \in D$, one needs to perturb each component v_i separately; while this can be done by varying only the flow utilities in the first-best problem (SA–F), it generally cannot be done in the same manner in the second-best problem due to the presence of the IC constraints.

vectors as rows has full rank (of at most $d-i$).[26] Below, Lemma H.1 shows that D3 holds in two leading cases: (i) if only "local" incentive constraints ((RIC$^*_{ji}$) with $j=i+1$) bind, it holds for all utilities and transition matrices, and (ii) regardless of which constraints bind, it holds for CARA utility and generic transition matrices.

- D4 is an asymptotic condition requiring that the subgradients of $P(\mathbf{v}^{(t)}, s^{(t)})$ do not explode "too quickly" as $t \to \infty$ under the optimal contract.[27] Unfortunately, D4 is difficult to directly verify in our setting. At the end of this SA, we describe how this difficulty can be bypassed in the i.i.d. special case.

Overall, we conclude that RZS's D1 and D2 always hold, D3 holds under mild conditions, and that the main technical barrier to establishing Condition R.5 in general is verifying RZS's D4.[28] We note that D1–D4 are merely *sufficient* for Condition R.5. Indeed, we conjecture that Condition R.5 is implied by Regularity (Conditions R.1–R.3), and that this can be shown via the sequential problem (SP) by adapting Morand and Reffett (2015, Theorem 3). We leave further study of these issues as an important task for future work.

Sufficient Conditions for D3. Let $\mathcal{M} := \Delta(S)^S$ denote the set of transition matrices $\mathbf{F} := [\mathbf{f}_i]_{i=1}^d$ on $S = \{1, \ldots, d\}$. Let $\mathcal{M}^\circ \subset \mathcal{M}$ denote the subset of fully connected transition matrices (i.e., those consistent with Assumption Markov). Note that, being the d-fold product of a $(d-1)$-dimensional simplex, $\mathcal{M}$ (respectively, $\mathcal{M}^\circ$) is a $(d-1)d$-dimensional compact (respectively, open) convex set. We denote the $(d-1)d$-dimensional Lebesgue measure as $\mathrm{Leb}_{(d-1)d}(\cdot)$. We have $\mathcal{L} := \mathrm{Leb}_{(d-1)d}(\mathcal{M}) = \mathrm{Leb}_{(d-1)d}(\mathcal{M}^\circ) > 0$.

LEMMA H.1: *For every $d \geq 2$, the following hold:*

(a) *For any $\mathbf{F} \in \mathcal{M}^\circ$ and utility function U satisfying Assumption DARA: If only local IC constraints bind,*[29] *then D3 holds.*

(b) *For any CARA utility function: There exists an $M \subseteq \mathcal{M}^\circ$ that is open, dense, and has full measure such that, regardless of which IC constraints bind, D3 holds for all $\mathbf{F} \in M$.*

The two parts of Lemma H.1 are complementary. Part (a) shows that, when the FOA is valid, D3 holds for all $\mathbf{F}$ and U. This covers all instances of the model with $d=2$, and also permits i.i.d. type process. Part (b) allows for any combination of binding IC constraints, but restricts attention to CARA utility and generic $\mathbf{F}$; when global IC constraints are binding, the genericity condition elides i.i.d. processes.

PROOF OF LEMMA H.1: For part (a), let $\mathbf{v} \in D$ and $i \in S\backslash\{d\}$ be given. By hypothesis, we have $g^j(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) > 0$ for all $j > i+1$. If $g^{i+1}(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) > 0$, there is nothing to prove. If $g^{i+1}(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) = 0$, then the full-rank condition is violated iff $\nabla_2 g^{i+1}(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) = \mathbf{0}$. By (H.1), the latter condition holds iff $\mathbf{f}_{i+1} = \psi'(v_i - \alpha\,\mathsf{E}^{\mathbf{f}_i}[\mathbf{w}_i], i+1, i)\mathbf{f}_i$, which is impossible because $\mathbf{f}_{i+1}, \mathbf{f}_i \in \Delta(S)$ and $\psi'(\cdot, i+1, i) < 1$ (recall Footnote 19). Thus, D3 holds.

For part (b), let $U(c) \equiv -e^{-\rho c}$ for some $\rho > 0$. For each $i \in S$, define $\theta_i := e^{-\rho\omega_i}$ so that, for all $j \geq i+1$, $\psi(u, j, i) \equiv \frac{\theta_j}{\theta_i} u$ and hence (H.1) becomes $\nabla_2 g^j(\cdot, \cdot, i) \equiv \alpha\{-\mathbf{f}_j + \frac{\theta_j}{\theta_i}\mathbf{f}_i\}$. For each

[26] For each $i \in S$, there are in total $d-i$ (RIC$^*_{ji}$) constraints with $j > i$, not all of which necessarily bind.

[27] Formally, D4 requires that that, for each $(\mathbf{v}^{(0)}, s^{(0)}) \in D \times S$, there exists a constant $B_0 \in \mathbb{R}$ such that $\lim_{t\to\infty} \mathsf{E}\big[\alpha^t G^{(t)} \zeta^{(t)} \mid (\mathbf{v}^{(0)}, s^{(0)})\big] = B_0$ for every measurable selection of subgradients $\zeta^{(t)} \in \partial P(\mathbf{v}^{(t)}, s^{(t)})$, where each $G^{(t)}$ is a (random) $d \times d$ matrix formed from pseudo-inverses of the matrices described in D3.

[28] We do not know of any examples in the literature where D1–D3 all hold but D4 fails. As discussed in RZS, known non-differentiabilities in dynamic contracting models are associated with violations of D1–D3.

[29] That is, if $g^j(\mathbf{v}, \mathbf{w}_i(\mathbf{v}), i) > 0$ for all $i, j \in S$ such that $j > i+1$ and $\mathbf{v} \in D$.

$i \in S\backslash\{d\}$ and $\mathbf{F} \in \mathcal{M}^\circ$, define the $(d-i)\times d$ matrix $B_i(\mathbf{F}) := [\mathbf{f}_j - \frac{\theta_j}{\theta_i}\mathbf{f}_i]_{j=i+1}^d$. If $B_i(\mathbf{F})$ has full (row) rank for every $i \in S\backslash\{d\}$, then D3 holds (regardless of which IC constraints bind). We claim that every $B_i(\mathbf{F})$ has full rank for all $\mathbf{F}$ in an open, dense, and full-measure subset $M \subseteq \mathcal{M}^\circ$. To this end, let $i \in S\backslash\{d\}$ be given. Let K_i be the (finite) index set of all the $(d-i)\times(d-i)$ submatrices of $B_i(\mathbf{F})$, with typical index $k \in K_i$ and corresponding submatrix $B_{ik}(\mathbf{F})$. For each $k \in K_i$, let $M_{ik} := \{\mathbf{F} \in \mathcal{M}^\circ : B_{ik}(\mathbf{F})$ has full rank$\}$. Since $B_i(\mathbf{F})$ has full rank iff there exists a $k \in K_i$ for which $B_{ik}(\mathbf{F})$ has full rank, we have $\{\mathbf{F} \in \mathcal{M}^\circ : B_i(\mathbf{F})$ has full rank$\} = \bigcup_{k\in K_i} M_{ik} =: N_i$. For each $k \in K_i$, note that $\mathbf{F} \in M_{ik}$ iff the determinant $\det(B_{ik}(\mathbf{F})) \neq 0$; since the mapping $\mathbf{F} \mapsto \det(B_{ik}(\mathbf{F}))$ is a non-constant polynomial on $\mathcal{M}^\circ$, $M_{ik} \subset \mathcal{M}^\circ$ is open and has full measure, i.e., $\text{Leb}_{(d-1)d}(M_{ik}) = \mathcal{L}$.[30] Thus, $N_i \subset \mathcal{M}^\circ$ is open and has full measure $\text{Leb}_{(d-1)d}(N_i) = \mathcal{L}$. Since $i \in S\backslash\{d\}$ was arbitrary, $M := \bigcap_{i=1}^{d-1} N_i = \{\mathbf{F} \in \mathcal{M}^\circ : B_i(\mathbf{F})$ has full rank $\forall i \in S\backslash\{d\}\}$ is open and has full measure $\text{Leb}_{(d-1)d}(M) = \mathcal{L}$; hence, M is also dense in $\mathcal{M}^\circ$.[31] This proves the claim. *Q.E.D.*

Verifying Condition R.5 in the i.i.d. Case. Let $\boldsymbol{\pi} \in \Delta(S)$ denote the (type-independent) transition probabilities. Recall the efficiency problem (Eff_i) from SA–C.3.1. As noted in Lemma C.7, this problem has (type-independent) value function $K : \mathcal{U} \to \mathbb{R}$, which represents the restriction of the (type-independent) value function $P : D \to \mathbb{R}$ to the (type-independent) efficient set $E \subsetneq D$, which is parameterized by (one-dimensional) *ex ante* promised utility $v \in \mathcal{U}$.[32] Formally, Lemma C.6(a)–(b) implies that (i) there is a bijection between promised utility vectors $\mathbf{v} \in E$ and their corresponding *ex ante* promised utilities $v = \mathbf{E}^{\boldsymbol{\pi}}[\mathbf{v}] \in \mathcal{U}$, and (ii) $K(v) = \min\{P(\mathbf{v}) : \mathbf{v} \in D \text{ s.t. } \mathbf{E}^{\boldsymbol{\pi}}[\mathbf{v}] = v\}$ for all $v \in \mathcal{U}$, where the minimum is attained on E.[33] Moreover, the optimal contract always maps to E: in the notation of SA–C, $\xi^c(\mathbf{v}, s, i) \in E$ for all $\mathbf{v} \in D$ and $i, s \in S$.

We claim that, if only local IC constraints bind, then K and P are both $\mathbf{C}^1$. We sketch the proof below; it sidesteps the need to directly check RZS's D4 (cf. their Corollary 3.1).

To begin, note that the above discussion implies that P and K satisfy

$$\begin{aligned} P(\mathbf{v}) = &\min_{(u_i,w_i)_{i\in S}\in(\mathcal{U}\times\mathcal{U})^d} \sum_{i\in S} \pi_i\,[C(u_i,i) + \alpha K(w_i)] \\ \text{s.t.} \quad & v_i = u_i + \alpha w_i \qquad \forall i \in S, \\ & v_i \geq \psi(u_{i-1}, i, i-1) + \alpha w_{i-1} \quad \forall i \in S, \end{aligned} \tag{H.2}$$

where $w_i := \mathbf{E}^{\boldsymbol{\pi}}[\mathbf{w}_i] \in \mathcal{U}$ is the *ex ante* continuation utility following report $i \in S$. Then, because $K(v) = \min\{P(\mathbf{v}) : \mathbf{v} \in D \text{ s.t. } \mathbf{E}^{\boldsymbol{\pi}}[\mathbf{v}] = v\}$ for all $v \in \mathcal{U}$, K satisfies

$$\begin{aligned} K(v) = &\min_{(u_i,w_i)_{i\in S}\in(\mathcal{U}\times\mathcal{U})^d} \sum_{i\in S} \pi_i\,[C(u_i,i) + \alpha K(w_i)] \\ \text{s.t.} \quad & v = \sum_{i\in S} \pi_i [u_i + \alpha w_i], \\ & u_i + \alpha w_i \geq \psi(u_{i-1}, i, i-1) + \alpha w_{i-1} \quad \forall i \in S. \end{aligned} \tag{H.3}$$

[30] It is a standard fact that, for any open connected set $U \subseteq \mathbb{R}^{(d-1)d}$ and non-constant polynomial $f : U \to \mathbb{R}$, the zero set $Z(f) := \{x \in U : f(x) = 0\}$ is closed (in U) and satisfies $\text{Leb}_{(d-1)d}(Z(f)) = 0$. Applying this fact to $U = \mathcal{M}^\circ$ and $f(\cdot) = \det(B_{ik}(\cdot))$ and taking complements yields the desired conclusion.

[31] By a standard argument, every open subset of $\mathcal{M}^\circ$ with full measure is dense.

[32] Note that we implicitly redefine the domains of K and P to reflect the fact that, in the i.i.d. case, they do not depend on $i \in S$. This minor abuse of notation simplifies the subsequent presentation.

[33] In Lemma C.6, Condition R.5 is only used to show that $K \in \mathbf{C}^1(\mathcal{U})$ in part (c).

This is the Bellman equation from Thomas and Worrall (1990), which features a single *ex ante* promise keeping constraint (and we include only the local downward IC constraints). Thus, $K \in \mathbf{C}^1(\mathcal{U})$ by a standard application of Benveniste and Scheinkman (1979).[34]

Next, we use this fact to show via (H.2) that $P \in \mathbf{C}^1(D)$. In particular, in (H.2) we can use the promise keeping constraints to solve out for $(u_i)_{i\in S}$, yielding a minimization problem over $(w_i)_{i\in S} \in \mathcal{U}^d$ subject only to the reduced-form IC constraints

$$h^i(\mathbf{v}, w_1, \dots, w_d) := v_i - \psi(v_{i-1} - \alpha w_{i-1}, i, i-1) - \alpha w_{i-1} \geq 0 \quad \forall i \in S.$$

An argument analogous to the proof of Lemma H.1(i) shows that the derivatives of the h^i functions with respect to $(w_i)_{i\in S}$ are linearly independent. Thus, RZS's D1–D3 hold for this reduced version of (H.2). But since $K \in \mathbf{C}^1(\mathcal{U})$, this suffices to show that $P \in \mathbf{C}^1(D)$ (cf. the special case of RZS's Proposition 3.1 in which the continuation value function is known to be smooth). This proves the claim.

REFERENCES

BENVENISTE, LAWRENCE M. AND JOSE A. SCHEINKMAN (1979): "On the Differentiability of the Value Function in Dynamic Models of Economics," *Econometrica*, 41, 727–733. [24, 26, 29]

BLOEDEL, ALEXANDER W., R. VIJAY KRISHNA, AND OKSANA LEUKHINA (2024): "Extended Supplement to 'Insurance and Inequality with Persistent Private Information'," . [1, 5, 24, 25]

GOLOSOV, MIKHAIL, ALEH TSYVINSKI, AND NICOLAS WERQUIN (2016): "Recursive Contracts and Endogenously Incomplete Markets," in *Handbook of Macroeconomics, Volume 2*. [29]

HATCHER, ALLEN (2001): *Algebraic Topology*, Cambridge University Press. [4]

LUENBERGER, DAVID G. (1969): *Optimization by Vector Space Methods*, Wiley. [1, 9]

MILGROM, PAUL AND ILYA SEGAL (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70 (2), 583–601. [5]

MORAND, OLIVIER AND KEVIN REFFETT (2015): "Lagrange multipliers in convex programs with applications to classical and nonoptimal stochastic one-sector growth models," *University of Connecticut and Arizona State University*. [27]

NORRIS, JAMES R. (1997): *Markov Chains*, Cambridge, UK: Cambridge University Press. [26]

RINCÓN-ZAPATERO, JUAN PABLO AND MANUEL S. SANTOS (2009): "Differentiability of the Value Function without Interiority Assumptions," *Journal of Economic Theory*, 144 (5), 1948–1964. [26]

ROCKAFELLAR, R. TYRELL (1970): *Convex Analysis*, Princeton, NJ: Princeton University Press. [9]

SHIRYAEV, ALBERT N. (1995): *Probability*, Springer. [13, 25]

THOMAS, JONATHAN P. AND TIM WORRALL (1990): "Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem," *Journal of Economic Theory*, 51, 367–390. [6, 29]

Co-editor [Name Surname; will be inserted later] handled this manuscript.

[34]Formally, suppose the menu $(u_i, w_i)_{i\in S}$ solves (H.3) at $v \in \mathcal{U}$. For $t \in \mathbb{R}$, define $x_d(t) := t$ and, for all $i < d$, inductively define $x_i(t) := \psi^{-1}(\psi(u_i, i+1, i) + x_{i+1}(t), i+1, i) - u_i$. (Since $\mathcal{U}$ is open, there is an $\varepsilon > 0$ such that every $u_i + x_i(t) \in \mathcal{U}$ for $|t| < \varepsilon$.) Then $(u_i + x_i(t), w_i)_{i\in S}$ is feasible in (H.3) at $v + g(t)$, where $g(t) := \sum_{i\in S} \pi_i x_i(t)$. (By construction, this is the essentially unique way to perturb the initial flow utilities without changing the amount of slack in any of the local downward IC constraints.) Each $x_i(\cdot)$ is $\mathbf{C}^1$ with $x_i'(\cdot) > 0$, and satisfies $x_i(0) = 0$. Then, since each $C(\cdot, i) \in \mathbf{C}^1(\mathcal{U})$, Benveniste and Scheinkman (1979) delivers $K \in \mathbf{C}^1(\mathcal{U})$. Thomas and Worrall (1990, Proposition 1) describe this argument informally; for CARA utility, it coincides with the construction in Golosov, Tsyvinski, and Werquin (2016, Lemma 1).