Online Appendix to "Optimal Long-term Health Insurance Contracts: Characterization, Computation, and Welfare Effects"

Soheil Ghili, Ben Handel, Igal Hendel, and Michael D. Whinston

August 15, 2022

1 Appendix A: Characterization of Equilibrium Contracts with One-Sided Commitment

To recount some basics from the main text: We suppose that there are a total of T periods, $t = 1, ..., T$. The consumer's within-period utility function is $u(\cdot)$. It is strictly increasing and strictly concave. Health expenses in period t are denoted m_t. The consumer's health status in period t is λ_t, which determines his period-t expected medical expenses, $\mathbb{E}(m_t|\lambda_t)$. The consumer's income in period t is y_t. We assume that the consumer's utility function $u(\cdot)$ and income path $(y_1, ..., y_T)$ are known. The consumer also has a switching cost incurred whenever he changes insurers, equal to $\sigma \geq 0$. (We will establish our proposition for the general case of a nonnegative switching cost; Proposition 1 will follow as the special case of $\sigma = 0$.)

We denote by $\Lambda^t_{t'}$ the consumer's history of health statuses from period t to period t', $\Lambda^t_{t'} \equiv (\lambda_t, ..., \lambda_{t'})$. Similarly, the consumer's history of medical expense realizations from period t to period t' is $M^t_{t'} \equiv (m_t, ..., m_{t'})$. We will refer to $\Lambda^t_{t'}$ as the consumer's "continuation health history" starting at period t. At the start of period t, the probability of the continuation health history $\Lambda^t_{t'}$ being reached depends only on the consumer's health history at period t, Λ^1_t, which we refer to as the consumer's period-t "health state," and is given by $f(\Lambda^t_{t'}|\Lambda^1_t)$.[1] Finally, we denote by $\left\langle \Lambda^t_{t'}, \Lambda^{t'+1}_{t''} \right\rangle$ a health history constructed by putting together $\Lambda^t_{t'}$ and $\Lambda^{t'+1}_{t''}$.

1.1 Contracts

We are concerned with identifying optimal contracts that may be signed at each date and history. Since at the start of a period t the future depends only on the consumer's health state Λ^1_t, an optimal contract will depend only this, and not on previous medical expense

[1] In our empirical work we suppose that $f(\cdot|\cdot)$ is a second-order Markov process, generated by a transition process $\widehat{f}(\lambda_{t+1}|\lambda_{t-1}, \lambda_t)$. As such, we will then refer to $\Lambda_t = (\lambda_{t-1}, \lambda_t)$ as the consumer's period-t health state.

realizations. We therefore denote a contract signed with the consumer at health history Λ_t^1 by $c_{\Lambda_t^1}(\cdot)$.[2] The contract $c_{\Lambda_t^1}(\cdot)$ is a function that specifies the consumer's consumption level in each future period $t' \geq t$ for each possible continuation history $(\Lambda_{t'+1}^{t+1}, M_{t'}^t)$.[3] Thus, the consumption level specified by $c_{\Lambda_t^1}(\cdot)$ in period $t' \geq t$ can in general be written as $c_{\Lambda_t^1}(\Lambda_{t'+1}^{t+1}, M_{t'}^t)$.

It will be useful in what follows to consider contracts that would break even if subsidized by some amount. To this effect, we say that contract $c_{\Lambda_t^1}(\cdot)$ breaks even with subsidy $S \in \mathbb{R}$ if

$$\Sigma_{\tau=t}^{T} \delta^{\tau-t} \left([y_\tau - \mathbb{E}[m_\tau | \Lambda_t^1] - \mathbb{E}[c_{\Lambda_t^1}(\Lambda_{\tau+1}^{t+1}, M_\tau^t) | \Lambda_t^1] \right) = -S \tag{9}$$

We say that the contract is a "zero profit contract" if it breaks even with subsidy $S = 0$, and we denote the set of all contracts signed at Λ_t^1 that break even with subsidy S by $B^S(\Lambda_t^1)$.

The value to the consumer of contract $c_{\Lambda_t^1}(\cdot)$ starting at health state Λ_t^1 is denoted $V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot))$ and is defined as follows:

$$V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot)) = \Sigma_{\tau=t}^{T} \delta^{\tau-t} [\mathbb{E}[u(c_{\Lambda_t^1}(\Lambda_{\tau+1}^{t+1}, M_\tau^t)) | \Lambda_t^1] \tag{10}$$

For $t' > t$, we denote by $c_{\Lambda_t^1 | \Lambda_{t'}^{t+1}}(\cdot)$ a "sub-contract" of $c_{\Lambda_t^1}(\cdot)$ that is given by looking at the consumption levels implied by $c_{\Lambda_t^1}(\cdot)$ (weakly) after the realization of continuation health history $\Lambda_{t'}^{t+1}$. Mathematically, $c_{\Lambda_t^1 | \Lambda_{t'}^{t+1}}(\cdot)$ could also be looked at as a stand-alone contract signed at the beginning of year t' given health state $\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle$. Obviously, $c_{\Lambda_t^1 | \Lambda_{t'}^{t+1}}(\cdot)$ being zero-profit neither implies nor is implied by $c_{\Lambda_t^1}(\cdot)$ being zero-profit.

Definition 1, repeated here, then describes an optimal contract given an initial subsidy level S:

Definition 3 $c_{\Lambda_t^1}^*(\cdot | S_t)$ *is an optimal contract signed in period t at health state Λ_t^1 with subsidy S_t if it solves the following maximization problem:*

$$\max_{c_{\Lambda_t^1}(\cdot) \in B^{S_t}(\Lambda_t^1)} V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot)) \tag{11}$$

$$s.t. \;\; V_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(c_{\Lambda_t^1 | \Lambda_{t'}^{t+1}}(\cdot)) \geq V_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(c^*_{\langle \Lambda_t^1, \Lambda_{t'}^{t+1} \rangle}(\cdot | -\sigma)) \text{ for all } \Lambda_{t'}^{t+1} \text{ with } t' > t$$

[2] In this appendix, we suppress the dependence of the contract on the consumer's type θ, consisting of his utility function $u(\cdot)$ and income path $y = (y_1 ..., y_T)$.

[3] Recall that λ_{t+1} and m_t are realized during period t and the consumption specified for period t can depend on them.

In what follows, we will denote the special case of $c^*_{\Lambda^1_t}(\cdot|0)$ (the optimal zero-profit contract) by $c^*_{\Lambda^1_t}(\cdot)$ for simplicity. Also, $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S)$ is the subcontract of optimal contract $c^*_{\Lambda^1_t}(\cdot|S)$ that starts in period t' at history $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$; $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot)$ is the special case of a subcontract of zero-profit optimal contract $c^*_{\Lambda^1_t}(\cdot)$.

Note that equation (11) provides a recursive definition of the optimal contract. The constraint in this definition makes sure that at no continuation health history $\Lambda^{t+1}_{t'}$ does the customer prefer to lapse to $c^*_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(\cdot|-\sigma)$, the optimal contract starting at health state $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$ with non-positive subsidy $-\sigma \leq 0$. The non-positive subsidy comes from the fact that an insurer seeking to lure the consumer away from the contract $c^*_{\Lambda^1_t}(\cdot|S)$ must effectively compensate the consumer for the fact that he incurs the switching cost σ.[4] The constraint ensures us that, following the realization of continuation health history $\Lambda^t_{t'}$, the consumer does not prefer to lapse to any other contract $c_{(\Lambda^1_t, \Lambda^{t+1}_{t'})}(\cdot)$ that would at least break even given the need to compensate the consumer for his switching cost, and that also satisfies no-lapsation.

To begin, we first prove a lemma demonstrating that an optimal contract signed in a period t always specifies at each period $t' \geq t$ and continuation health history $\Lambda^t_{t'}$ a deterministic consumption level; that is, consumption that does not depend upon the realization during period t' of the consumer's period $t'+1$ health status, nor the consumer's continuation medical expenses from period t to t', $M^t_{t'}$. In particular, upon arriving at any period t' and continuation health history $\Lambda^t_{t'}$, the contract offers the consumer full within-period insurance against his period t' medical expenses.

Lemma 2 *For any $t' \geq t$ and $(\Lambda^1_t, \Lambda^{t+1}_{t'+1}, M^t_{t'}, S)$, we have:*

$$c^*_{\Lambda^1_t}(\Lambda^{t+1}_{t'+1}, M^t_{t'}|S) = c^*_{\Lambda^1_t}(\Lambda^{t+1}_{t'}|S).$$

Proof of Lemma 2. Consider a period t' and two continuation histories $(\overline{\Lambda}^{t+1}_{t'+1}, \overline{M}^t_{t'}) \neq (\widehat{\Lambda}^{t+1}_{t'+1}, \widehat{M}^t_{t'})$ with $\overline{\Lambda}^{t+1}_{t'} = \widehat{\Lambda}^{t+1}_{t'} \equiv \Lambda^{t+1}_{t'}$ that can both happen with positive probability conditional on Λ^1_t, and suppose that, contrary to the statement of the lemma,

$$c^*_{\Lambda^1_t}(\overline{\Lambda}^{t+1}_{t'+1}, \overline{M}^t_{t'}|S) \neq c^*_{\Lambda^1_t}(\widehat{\Lambda}^{t+1}_{t'+1}, \widehat{M}^t_{t'}|S).$$

We show that one could then construct a contract that is strictly preferred by the customer to $c^*_{\Lambda^1_t}(\cdot|S)$ and does not violate no-lapsation or the budget constraint. To do this, we

[4] The consumption level specified in the contract offered by the new insurer is *net* of the consumer's switching cost.

consider contract $c_{\Lambda_t^1}(\cdot)$ such that for $(\Lambda_{t'+1}^{t+1}, M_{t'}^t) \in \{(\overline{\Lambda}_{t'+1}^{t+1}, \overline{M}_{t'}^t), (\widehat{\Lambda}_{t'+1}^{t+1}, \widehat{M}_{t'}^t)\}$ we have

$$c_{\Lambda_t^1}(\Lambda_{t'+1}^{t+1}, M_{t'}^t|S) = \mathbb{E}[c^*_{\Lambda_t^1}(\Lambda_{t'+1}^{t+1}, M_{t'}^t|S)|\ (\langle \Lambda_t^1, \Lambda_{t'+1}^{t+1}\rangle, M_{t'}^t) \in \{(\left\langle \Lambda_t^1, \overline{\Lambda}_{t'+1}^{t+1}\right\rangle, \overline{M}_{t'}^t), (\left\langle \Lambda_t^1, \widehat{\Lambda}_{t'+1}^{t+1}\right\rangle, \widehat{M}_{t'}^t)\}]$$

and otherwise,

$$c_{\Lambda_t^1}(\Lambda_{t'+1}^{t+1}, M_{t'}^t|S) = c^*_{\Lambda_t^1}(\Lambda_{t'+1}^{t+1}, M_{t'}^t|S).$$

Given that the utility function $u(\cdot)$ is strictly concave, this consumption-smoothing modification will imply that the customer strictly prefers $c_{\Lambda_t^1}(\cdot)$ over $c^*_{\Lambda_t^1}(\cdot|S)$. Also, this modification does not change the contract's expected profit. Finally, this modification weakly improves the expected utility of the contract at all possible super-histories of Λ_t^1. Thus, contract $c_{\Lambda_t^1}(\cdot)$ also satisfies no-lapsation. But this contradicts the optimality of $c^*_{\Lambda_t^1}(\cdot|S)$.■

Given lemma (2), we can simplify notation and write contracts in the form of $c_{\Lambda_t^1}(\Lambda_{t'}^{t+1}|S)$. However, in what follows it will actually be more convenient and clearer (despite some redundancy in the notation) to write the contract as a function of the full health history $\Lambda_{t'}^1 = \left\langle \Lambda_t^1, \Lambda_{t'}^{t+1}\right\rangle$ that has been reached at date $t' \geq t$; hence in the form of $c_{\Lambda_t^1}(\Lambda_{t'}^1|S)$.

We introduce two more notations on comparing contracts to one another before we turn to the proposition and its proof. First, for two contracts $c_{\Lambda_t^1}(\cdot)$ and $\hat{c}_{\Lambda_t^1}(\cdot)$ offered at the same health state Λ_t^1, we say the former is "preferred" to the latter, and write $c_{\Lambda_t^1}(\cdot) \succeq \hat{c}_{\Lambda_t^1}(\cdot)$ if $V_{\Lambda_t^1}(c_{\Lambda_t^1}(\cdot)) \geq V_{\Lambda_t^1}(\hat{c}_{\Lambda_t^1}(\cdot))$.

Second, for two contracts signed at the same health state Λ_t^1, we say $c_{\Lambda_t^1}(\cdot)$ "dominates" $\hat{c}_{\Lambda_t^1}(\cdot)$, and write $c_{\Lambda_t^1}(\cdot) \geq \hat{c}_{\Lambda_t^1}(\cdot)$, if the former offers a weakly higher consumption level than the latter at any possible future health history, including period t. That is, for every $t' \geq t$ and history $\Lambda_{t'}^1 = \left\langle \Lambda_t^1, \Lambda_{t'}^{t+1}\right\rangle$ with $f(\Lambda_{t'}^{t+1}|\Lambda_t^1) > 0$, we have $c_{\Lambda_t^1}(\Lambda_{t'}^1) \geq \hat{c}_{\Lambda_t^1}(\Lambda_{t'}^1)$. Note that if $c_{\Lambda_t^1}(\cdot) \geq \hat{c}_{\Lambda_t^1}(\cdot)$, then for any $t' > t$ and $\Lambda_{t'}^{t+1}$ we have $c_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot) \geq \hat{c}_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot)$.[5]

The strict versions of the above two relationships (i.e. $\succ$ and $>$) are defined in the natural way.

1.2 Proposition and Proof

We establish the following result, from which Proposition 1 follows as the special case where the switching cost σ equals zero.[6]

Proposition 4 *The optimal contract $c^*_{\Lambda_t^1}(\cdot)$ is fully characterized by the zero-profit condition and, for all $t' > t$ and $\Lambda_{t'}^t$ such that $f(\Lambda_{t'}^t|\Lambda_t^1) > 0$, the condition that the consumer receives*

[5] The same need not be true for $\succeq$.

[6] The proof in this subsection assumes the consumer does not engage in secret savings; we establish this fact formally in Appendix B.

the following certain consumption level:

$$c^*_{\Lambda^1_t}(\Lambda^1_{t'}) = \max\{c^*_{\Lambda^1_t}(\Lambda^1_t), \max_{\tau\in\{t+1,\ldots,t'\}} c^*_{\langle\Lambda^1_t,\Lambda^{t+1}_\tau\rangle}(\langle\Lambda^1_t,\Lambda^{t+1}_\tau\rangle)| - \sigma)\}. \tag{12}$$

In words, the optimal contract $c^*_{\Lambda^1_t}(\cdot)$ offers in each period $t' > t$ at history $\Lambda^1_{t'} = \langle\Lambda^1_t, \Lambda^{t+1}_{t'}\rangle$ the maximum among the first-period consumption levels offered by all the equilibrium contracts available along the way on continuation health history $\Lambda^{t+1}_{t'}$.

The proof strategy is based on strong induction: We assume the proposition is true for the optimal contracts $c^*_{\Lambda^1_{t'}}(\cdot)$ at all $\Lambda^1_{t'}$ with $t' > t$, and then show it is also true for the period-t optimal contracts $c^*_{\Lambda^1_t}(\cdot)$ for any Λ^1_t. To establish the result, we show that if for some Λ^1_t, optimal contract $c^*_{\Lambda^1_t}(\cdot)$ does not satisfy (12), then there is a "modification of" $c^*_{\Lambda^1_t}(\cdot)$ that (i) is strictly preferred to $c^*_{\Lambda^1_t}(\cdot)$ by the consumer; and (ii) satisfies no-lapsation and zero-profit.

Before we get to the proof itself, we introduce a notation on how to "modify" a contract.

Definition 4 *Let $\min\{t', t''\} \geq t$. We say contract $\hat{c}_{\Lambda^1_t}(\cdot)$ is an ε-transfer, from $\Lambda^1_{t'}$ to $\Lambda^1_{t''}$, on contract $c_{\Lambda^1_t}(\cdot)$, and write $\hat{c}_{\Lambda^1_t}(\cdot) = tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^1_{t'}, \Lambda^1_{t''}]$ if:*

1. $\hat{c}_{\Lambda^1_t}(\Lambda^1_{t'}) = c_{\Lambda^1_t}(\Lambda^1_{t'}) - \varepsilon$
2. $\hat{c}_{\Lambda^1_t}(\Lambda^1_{t''}) = c_{\Lambda^1_t}(\Lambda^1_{t''}) + \left[\varepsilon \times \frac{f(\Lambda^{t+1}_{t'}|\Lambda^1_t)}{f(\Lambda^{t+1}_{t''}|\Lambda^1_t)} \times \delta^{t'-t''}\right]$
3. *For all $\tau \geq t$ and $\Lambda^t_\tau \notin \{\Lambda^t_{t'}, \Lambda^t_{t''}\}$, we have $\hat{c}_{\Lambda^1_t}(\Lambda^t_\tau) = c_{\Lambda^1_t}(\Lambda^t_\tau)$*

In words, this ε-transfer just transfers some consumption between health histories $\Lambda^1_{t'}$ and $\Lambda^1_{t''}$ after applying a multiplier to the transfer to keep the discounted expected consumption unchanged. Our improvements on $c^*_{\Lambda^1_t}(\cdot)$ in the counter-positive strategy will be constructed using ε-transfers. We record two facts about such transfers:

Remark 1 *ε-transfers preserve the expected discounted profit: If $c^*_{\Lambda^1_t}(\cdot) \in B^S(\Lambda^1_t)$ for some $S \in \mathbb{R}$, then $tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^1_{t'}, \Lambda^1_{t''}] \in B^S(\Lambda^1_t)$.*

Remark 2 *For every $\Lambda^t_{t'}$ and $\Lambda^t_{t''}$ with $c_{\Lambda^1_t}(\Lambda^1_{t'}) > c_{\Lambda^1_t}(\Lambda^1_{t''})$ there exists an $\varepsilon_0 > 0$ such that for all $\varepsilon \leq \varepsilon_0$ we have $tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^t_{t'}, \Lambda^t_{t''}] \succ c_{\Lambda^1_t}(\cdot)$.*

Remark 3 follows immediately from the fact that the ε-transfer does not change the expected discounted consumption in the contract, while Remark 4 follows because of the consumer is strictly risk averse [$u(\cdot)$ is strictly concave].

Before proceeding to the proof of the proposition, we observe that in any optimal contract, the continuation contract specified at every future health history must itself be an optimal contract starting at that history for some subsidy:

Claim 1 *For $t' > t$, define $S_{t'}$ as the expected loss sustained by the insurer under contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ after the realization of health history $\Lambda^1_{t'} = \langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$. Formally:*

$$S_{t'} = \Sigma^T_{\tau=t'} \delta^{\tau-t'} \left(\mathbb{E}[c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \Lambda^{t+1}_\tau \rangle |S_t) - y_\tau - m_\tau | \Lambda^1_{t'}] \right).$$

Then, the following is true:

$$c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t) = c^*_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(\cdot|S_{t'}), \tag{13}$$

where

In words, Claim 2 states that any continuation contract $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t)$ of $c^*_{\Lambda^1_t}(\cdot|S_t)$ is in fact the optimal solution to the generalized problem outlined in Definition 1 for history $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$ when the subsidy available to the consumer is exactly the amount $S_{t'}$.

Proof of Claim 2. If at any continuation history $\Lambda^t_{t'}$ the condition in the claim did not hold we could replace the continuation contract $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t)$ by $c^*_{\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle}(\cdot|S_{t'})$ and do strictly better for the consumer without violating no-lapsation or changing the required subsidy S_t for contract $c^*_{\Lambda^1_t}(\cdot|S_t)$, a contradiction to the optimality of $c^*_{\Lambda^1_t}(\cdot|S_t)$. ■

We now turn to proving the proposition. To do so, we will actually prove a more general statement than the proposition, using strong induction on the number of periods. The following lemma is the general result that implies our proposition:

Lemma 3 *Consider optimal contract $c^*_{\Lambda^1_t}(\cdot|S_t)$. There exists a unique $\bar{c} \in \mathbb{R}$ such that $c^*_{\Lambda^1_t}(\Lambda^1_t|S_t) = \bar{c}$, and for any $t' > t$ and $\Lambda^{t+1}_{t'}$ such that $f(\Lambda^{t+1}_{t'}|\Lambda^1_t) > 0$, we have*

$$c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle |S_t) = \max\{\bar{c}, c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle | -\sigma)\}. \tag{14}$$

In words, Lemma 7 says that at any subsequent period t' and history $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$, contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ gives the larger value between (i) consumption that it immediately gives, and (ii) the consumption that the optimal, break-even contract with subsidy $-\sigma$ signed in the beginning of period $t+1$ at history $\langle \Lambda^1_t, \lambda_{t+1} \rangle$ would offer.

Note that condition (30) of the lemma implies that any two optimal contracts signed at time t and health history Λ^1_t, but with differing subsidies $S''_t > S'_t$, are ordered by the dominance relation according to the level of the inital consumptions they specify, which by the break-even condition are ordered according to the size of the subsidies; that is, $c^*_{\Lambda^1_t}(\cdot|S''_t) > c^*_{\Lambda^1_t}(\cdot|S'_t)$.

Proof of Lemma 7. The proof goes by induction. For $t = T$, the result is straightforward, given that there is no period after $t = T$; at that point, $c_{\Lambda_T^1}(\Lambda_T^1|S) = y_T - \mathbb{E}[m_T|\lambda_T] + S$. We now turn to the proof for $t < T$, assuming, by way of induction, that the result holds for any $\tau > t$ and any S_τ. We begin by showing that (30) holds for any period $t+1$ history $\langle \Lambda_t^1, \lambda_{t+1} \rangle$; i.e., that

$$c^*_{\Lambda_t^1|\lambda_{t+1}}(\langle \Lambda_t^1, \lambda_{t+1} \rangle |S_t) = \max\{c^*_{\Lambda_t^1}(\Lambda_t^1|S_t), c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\langle \Lambda_t^1, \lambda_{t+1} \rangle | -\sigma)\}. \tag{15}$$

To this end, we consider two cases regarding history nodes ending at period $t+1$.

Case 1. *At history $\langle \Lambda_t^1, \lambda_{t+1} \rangle$, the no-lapsation condition is binding for contract $c^*_{\Lambda_t^1}(\cdot|S_t)$.* Formally:

$$V_{\Lambda_t^1}(c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t)) = V_{\Lambda_t^1}(c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot| -\sigma)). \tag{16}$$

Note that by Claim 2, the continuation contract $c^*_{\Lambda_t^1|\Lambda_{t'}^{t+1}}(\cdot|S_t)$ is itself the optimal contract $c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot|S_{t+1})$ for some S_{t+1}. Thus, (32) implies that

$$c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t) = c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot| -\sigma). \tag{17}$$

Next, note that the immediate consumption $c^*_{\Lambda_t^1}(\Lambda_t^1|S_t)$ in contract $c^*_{\Lambda_t^1}(\cdot|S_t)$ must satisfy

$$c^*_{\Lambda_t^1}(\Lambda_t^1|S_t) \le c^*_{\Lambda_t^1}(\langle \Lambda_t^1, \lambda_{t+1} \rangle |S_t) = c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\langle \Lambda_t^1, \lambda_{t+1} \rangle | -\sigma), \tag{18}$$

otherwise an ε-transfer from the immediate consumption at history Λ_t^1 to history $\langle \Lambda_t^1, \lambda_{t+1} \rangle$ would strictly improve the contract $c^*_{\Lambda_t^1}(\cdot|S_t)$ from the perspective of the consumer, without changing the expected profit. This transfer would also satisfy no-lapsation, given that it weakly increases the consumption given by $c^*_{\Lambda_t^1}(\cdot|S_t)$ at any history that happens strictly after time t. Thus, (31) holds at all period-$(t+1)$ histories at which lapsation binds.

Case 2. *At history $\langle \Lambda_t^1, \lambda_{t+1} \rangle$, the no-lapsation condition is not binding for contract $c^*_{\Lambda_t^1}(\cdot|S_t)$.* That is,

$$V_{\Lambda_t^1}(c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t)) > V_{\Lambda_t^1}(c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot| -\sigma)). \tag{19}$$

As in the previous case, given Claim 2, the continuation contract $c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t)$ is itself the optimal contract $c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\cdot|S_{t+1})$ for some S_{t+1}. Therefore, by our induction assumption, there is some $\bar{c}$ such that (i) $\bar{c} = c^*_{\Lambda_t^1|\lambda_{t+1}}(\langle \Lambda_t^1, \lambda_{t+1} \rangle |S_t)$ and (ii) for any history $\langle \Delta_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \rangle$ with $t' \ge t+2$ we have:

$$c^*_{\Lambda_t^1|\lambda_{t+1}}(\langle \Delta_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \rangle |S_t) = \max\{\bar{c}, c^*_{\langle \Lambda_t^1, \lambda_{t+1} \rangle}(\langle \Delta_t^1, \lambda_{t+1}, \Lambda_{t'}^{t+2} \rangle | -\sigma)\}. \tag{20}$$

But this, combined with inequality (36), tells us it must be that

$$\bar{c} > c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| - \sigma). \tag{21}$$

We now claim that

$$\bar{c} = c^*_{\Lambda^1_t}(\Lambda^1_t | S_t). \tag{22}$$

That is, $\bar{c}$ is equal to the immediate consumption offered by contract $c^*_{\Lambda^1_t}(\cdot|S_t)$. To see this, note that if $\bar{c} > c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)$, then an ε-transfer from the history $(\Lambda^1_t, \lambda_{t+1})$ to the immediate history Λ^1_t will increase the consumer's expected utility, will not change the expected profit from the contracts, and will preserve no-lapsation if ε is small enough, given (36). This contradicts the assumption that $c^*_{\Lambda^1_t}(\cdot|S_t)$ is the optimal contract. Conversely, if $\bar{c} < c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)$, the reverse ε-transfer will strictly increase the consumer's expected utility and preserve the insurer's expected profit. It also preserves no-lapsation since it weakly increases consumption at any history strictly after Λ^1_t. Thus, (31) also holds at all period-$(t+1)$ histories at which lapsation does not bind.

To sum up, cases 1 and 2 show that for *any* history λ_{t+1}, no matter whether no-lapsation is binding or not, (31) holds. Next, we combine (31) with the induction assumption to extend the argument, which currently applies only to period-$(t+1)$ histories $\langle \Lambda^1_t, \lambda_{t+1} \rangle$, also to any history $\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle$ with $t' > t+1$. By Claim 2, we know that for some appropriate S_{t+1}, we have:

$$c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| S_t) = c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| S_{t+1}) \tag{23}$$

and

$$c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| S_t) = c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| S_{t+1}) \tag{24}$$

By induction, we know that

$$c^*_{(\Lambda^1_t, \lambda_{t+1})}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| S_{t+1}) = \max\{c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| S_{t+1}), c^*_{\langle \Lambda^1_t, \lambda_{t+1}, \lambda_{t+2} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| -\sigma)\} \tag{25}$$

Replacing into equation (42) from (40) and (41), we get:

$$c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| S_t) = \max\{c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| S_t), c^*_{\langle \Lambda^1_t, \lambda_{t+1}, \lambda_{t+2} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| - \sigma)\} \tag{26}$$

Now, substituting for $c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| S_t)$ from (31), we get:

$$
\begin{aligned}
& c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| S_t) & (27) \\
= \;& \max\{\max\{c^*_{\Lambda^1_t}(\Lambda^1_t | S_t), c^*_{(\Lambda^1_t, \lambda_{t+1})}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| -\sigma)\}, c^*_{\langle \Lambda^1_t, \lambda_{t+1}, \lambda_{t+2} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| -\sigma)\} \\
= \;& \max\{c^*_{\Lambda^1_t}(\Lambda^1_t | S_t), \max\{c^*_{(\Lambda^1_t, \lambda_{t+1})}(\langle \Lambda^1_t, \lambda_{t+1} \rangle \,| -\sigma), c^*_{\langle \Lambda^1_t, \lambda_{t+1}, \lambda_{t+2} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| -\sigma)\}\}
\end{aligned}
$$

But by our induction assumption, the inner maximum equals $c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \lambda_{t+1}, \Lambda^{t+2}_{t'} \rangle \,| -\sigma)$; substituting this into (44) tells us that (30) holds for contracts signed in period t. Applying induction establishes the lemma.■

Proof of Proposition 4. Applying Lemma 7 to the special case of $S_t = 0$, we get that for any Λ^1_t and $\Lambda^{t+1}_{t'}$ such that $f(\Lambda^{t+1}_{t'} | \Lambda^1_t) > 0$, we have that

$$
c^*_{\Lambda^1_t}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle) = \max\{c^*_{\Lambda^1_t}(\Lambda^1_t), c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle \,| -\sigma)\}
$$

Since Lemma 7 holds for $c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\cdot | -\sigma)$ as well, we can expand $c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \rangle \,| -\sigma)$ in the same way as above. Then we can do this again and again, until we get that for all $t' > t$ and $\Lambda^t_{t'}$ such that $f(\Lambda^t_{t'} | \Lambda^1_t) > 0$,

$$
c^*_{\Lambda^1_t}(\Lambda^1_{t'}) = \max\{c^*_{\Lambda^1_t}(\Lambda^1_t), \max_{\tau \in \{t+1, \ldots, t'\}} c^*_{\langle \Lambda^1_t, \Lambda^{t+1}_\tau \rangle}(\Lambda^1_\tau) | -\sigma)\}, \tag{28}
$$

which is exactly the statement made in the proposition.■

2 Appendix B: Self-Selection

In this appendix we prove Proposition 2. We first establish the following Lemma:

Lemma 4 *Let $p = (p_\tau, \ldots, p_T)$ and $\widehat{p} = (\widehat{p}_\tau, \ldots, \widehat{p}_T)$ be guaranteed premium paths (that start in period τ) such that $\widehat{p} \geq p$ and consider a type θ consumer who is in health state Λ^1_τ in period τ. Suppose that (i) the insurer earns a non-negative expected payoff when guaranteed premium path p is chosen by the consumer (given the consumer's subsequent optimal lapsation behavior), (ii) under premium path p, in every period $t > \tau$ and health state Λ^1_t in which the consumer optimally does not lapse, the insurer's expected continuation payoff is non-positive, and (iii) the consumer never secretly saves when facing either of these premium paths. Then the insurer's expected continuation payoff is non-negative if premium path $\widehat{p}$ is chosen in period τ by the consumer.*

Proof. Let $U(t, \Lambda_t^1)$ and $\widehat{U}(t, \Lambda_t^1)$ denote the type θ consumer's continuation payoff in period t at health state Λ_t^1 given optimal lapsation behavior under p and $\widehat{p}$ respectively. Let $H_L(t)$ and $\widehat{H}_L(t)$ denote the sets of health states at which the consumer optimally lapses in period t, under p and $\widehat{p}$ respectively; $H_{NL}(t)$ and $\widehat{H}_{NL}(t)$ are the complementary sets of health states at which the consumer does not lapse. Finally, let $\Pi(t, \Lambda_t^1)$ and $\widehat{\Pi}(t, \Lambda_t^1)$ denote the insurer's expected continuation payoff at (t, Λ_t^1) given the consumer's optimal lapsation behavior under p and $\widehat{p}$ respectively. Assumption (i) therefore says that $\Pi(\tau, \Lambda_\tau^1) \geq 0$, while assumption (ii) says that $\Pi(t, \Lambda_t^1) \leq 0$ if $t > \tau$ and $\Lambda_t^1 \in H_{NL}(t)$ [of course, $\Pi(t, \Lambda_t^1) = 0$ for all $\Lambda_t^1 \in H_L(t)$].

Note, first, that $U(t, \Lambda_t^1) \geq \widehat{U}(t, \Lambda_t^1)$ for all (t, Λ_t^1): starting in period t, the consumer who faces p could adopt the same lapsation behavior as when facing $\widehat{p}$ and receive a weakly higher continuation payoff since under p he would be facing lower premia, and his optimal lapsation behavior under p yields a still higher payoff.[7] Next, the fact that $U(t, \Lambda_t^1) \geq \widehat{U}(t, \Lambda_t^1)$ for all (t, Λ_t^1) implies that $H_L(t) \subseteq \widehat{H}_L(t)$: in any health state in which the consumer lapses in period t when facing p, he also lapses when facing $\widehat{p}$. Finally, consider the expected payoff of the insurer starting at (τ, Λ_τ^1) under p. This is the probability weighted average of the payoffs achieved along the various possible sequences of health states $(\Lambda_\tau^1, ..., \Lambda_T^1)$. For each sequence the insurer earns premiums and incurs costs until the consumer lapses. Since $H_L(t) \subseteq \widehat{H}_L(t)$, each such sequence hits lapsation weakly earlier under $\widehat{p}$ than under p. Since, under path p, $\Pi(t, \Lambda_t^1) \leq 0$ if $t > \tau$ and $\Lambda_t^1 \in H_{NL}(t)$, the earlier termination behavior under $\widehat{H}_L$ (but earning the same premiums p prior to lapsation) would weakly raise the expected payoff earned by the insurer for the sequence by changing a non-positive expected continuation payoff into a continuation payoff of zero. Moreover, the fact that the premiums earned until lapsation are higher under $\widehat{p}$ than under p, while the expected costs are the same, means that a change from premium path p to path $\widehat{p}$, holding lapsation behavior fixed at $\widehat{H}_L$, would further raise the insurer's expected payoff earned from this health state sequence. As a result, $\Pi(\tau, \Lambda_\tau^1) \leq \widehat{\Pi}(\tau, \Lambda_\tau^1)$. ∎

We next establish the following Lemma:

Lemma 5 *Suppose that in each period $t \geq \tau$ the menu of contracts offered to a consumer who is in health state Λ_t^1 and wishes to sign a new contract is the set of optimal guaranteed premium path contracts for that consumer, $\{p_t^{\theta*}(\Lambda_t^1)\}_{\theta \in \Theta}$, and that moreover, in each period $t > \tau$ this menu is self-selective and induces no secret savings. Then a type θ consumer in health state Λ_τ^1 will not secretly save when facing guaranteed premium path $p_\tau^{\theta*}(\Lambda_\tau^1)$.*

[7] Note that the consumer who lapses in a period t when in some health state Λ_t^1 would receive the same new contract regardless of whether he was lapsing from p or from $\widehat{p}$.

Proof. Observe first that, under the assumptions of this lemma, if the consumer does not secretly save and then lapses in period $t > \tau$ when in health state Λ_t^1 his new insurance contract will have guaranteed premium path $p_t^{\theta *}(\Lambda_t^1) = p_\tau^{\theta *}(\Lambda_\tau^1) - \Delta(t, \Lambda_t^1)$ for some $\Delta(t, \Lambda_t^1) \in \mathbb{R}$.[8] Thus, he will optimally lapse in that period and state if and only if $\Delta(t, \Lambda_t^1) > 0$; that is, he lapses if and only if he gets a cheaper guaranteed premium path.

Next, we argue that if the consumer instead secretly saves under contract $p_\tau^{\theta *}(\Lambda_\tau^1)$, then he will optimally lapse whenever he would have if he did not secretly save (and possibly in additional states as well). To see this point, suppose that the consumer has secretly saved prior to arriving in period t in health state Λ_t^1 and he chooses not to lapse when he would have if he had not secretly saved. Then he would be better off instead lapsing and choosing the same new contract choice as if he had not secretly saved (choosing the cheaper guranteed premium contract that he would have lapsed to if he had not secretly saved), while keeping his future lapsation and savings behavior unchanged: doing so would only change his realized utility until the next lapsation, and would raise his payoff until that point in time by lowering his premiums.

Next, since in an optimal contract the insurer's continuation profits are always non-positive, by hastening lapsation secret savings can only weakly raise the profit of the insurer offering the consumer contract $p_\tau^{\theta *}(\Lambda_\tau^1)$. Moreover, the assumptions of the lemma imply that (because of self-selection and no future secret savings) all insurers providing the consumer with insurance after lapsation from contract $p_\tau^{\theta *}(\Lambda_\tau^1)$ will earn zero.[9] Thus, the total profit of insurers is non-negative with secret savings.

Finally, since total insurer profit is non-negative (and continuation profits are never strictly positive), the consumption path that results from secret savings was feasible in the no-secret savings problem. Hence, secret savings cannot raise the consumer's discounted expected utility and therefore the consumer will not prefer to secretly save. ■

We next establish the following Lemma:

Lemma 6 *Suppose that in each period $t \geq \tau$ the menu of contracts offered to a consumer who is in health state Λ_t^1 and wishes to sign a new contract is the set of optimal guaranteed premium path contracts, $\{p_t^{\theta *}(\Lambda_t^1)\}_{\theta \in \Theta}$, and that moreover, in each period $t > \tau$ this menu is self-selective and induces no secret savings. Then an insurer earns a non-negative continuation expected discounted profit if in period τ a type θ' consumer in health state Λ_τ^1 chooses the guaranteed premium path $p_\tau^{\theta *}(\Lambda_\tau^1)$ that is intended for a type θ consumer in health state Λ_τ^1.*

[8]That is, the new guaranteed premium path will differ from the old one by the same amount in each period (starting in period t).

[9]Note that a consumer with utility function u who arrives in period t in health state Λ_t^1 with savings S and who has remaining income path $y = (y_t, ..., y_T)$ is equivalent to a consumer who has income path $y^S \equiv (y_t + S, ..., y_T) \in \Theta$ and will self-select the policy intended for type $\theta = (y^S, u)$.

Proof. The proof is by induction. Consider the following induction hypothesis:

Induction Hypothesis: Under contract $p_\tau^{\theta *}(\Lambda_\tau^1)$, if starting in period $t > \tau$ the consumer has not yet lapsed, lapsation behavior of type θ' starting in period t is either (A) the same as for type θ (meaning, it is the same after any history of health states between periods τ and t and sequence of decisions not to lapse), or (B) different and raises the continuation expected discounted profit of the insurer starting in period t compared to the continuation payoff the insurer receives when facing a type θ consumer.

Observe that the Induction Hypothesis holds if $t = T$, since then lapsation behavior is the same for type θ' as for type θ – both types lapse if and only if $E[m_T|\lambda_T] < p_T^\theta$, where p_T^θ is the last period price in guaranteed premium path $p_\tau^{\theta *}(\Lambda_\tau^1) \equiv (p_\tau^\theta, ..., p_T^\theta)$.

Now suppose that the Induction Hypothesis holds for periods $t, ..., T$, and consider period $t-1$ ($\geq \tau + 1$) after some previous history of health states and a sequence of decisions in which the consumer has not yet lapsed. Recall that by Lemma 5 the type θ consumer does not secretly save; however, the type θ' consumer may.

Suppose first that, under $p_\tau^{\theta *}(\Lambda_\tau^1)$, period $t-1$ lapsation behavior is different for type θ' than for type θ in a health state Λ_{t-1}^1 in which type θ would not lapse. Then lapsation by type θ' in state Λ_{t-1}^1 would remove a continuation that had a weakly negative continuation payoff for the insurer when facing type θ and replace it with a zero continuation payoff when facing type θ'.

Suppose, instead, that state Λ_{t-1}^1 is one in which type θ would lapse in period $t-1$, choosing a contract with premium path $\widehat{p}$, while type θ' does not lapse. We will show that this changes what would have been a zero payoff continuation for the insurer into a continuation with a non-negative expected payoff when facing type θ'. By the self-selection assumption, we know that the contract $\widehat{p}$ that type θ chooses is $p_{t-1}^{\theta *}(\Lambda_{t-1}^1)$, the optimal guaranteed premium path contract for that consumer, so the insurer offering that contract breaks even. Note now that since that contract induces the type θ consumer to lapse there is a $\Delta > 0$ such that $\widehat{p}_k = p_k^\theta - \Delta$ for all periods $k \geq t-1$ (this is true because the two guaranteed premium paths differ only in offering different initial premiums and then the premium change each period equals the change in the type θ's income). Hence, by Lemma 4, if the type θ consumer were instead not to lapse from path $p_\tau^{\theta *}(\Lambda_\tau^1) \equiv (p_\tau^\theta, ..., p_T^\theta)$ in this state, the insurer's expected continuation payoff would be non-negative. But the Induction Hypothesis then implies that it is also non-negative (and hence at least as large as the continuation payoff of zero arising when facing type θ) when the type θ' consumer does not lapse in this state: the insurer's payoffs in period $t-1$ from the two types are the same as both the premium paid and the expected medical costs are the same for the two types. The transitions to the period t state Λ_t^1 are also the same. But, by the Induction Hypothesis, the insurer's expected continuation payoff under contract $p_\tau^{\theta *}(\Lambda_\tau^1)$ is weakly higher starting in period t when facing the type θ' consumer than when facing a type θ consumer. So the Induction Hypothesis holds in period $t-1$, and hence – applying induction – in period $\tau+1$.

Finally, consider period τ. The argument is similar to that above: If a type θ' consumer in health state Λ_τ^1 chooses the premium path $p_\tau^{\theta *}(\Lambda_\tau^1)$ intended for a type θ consumer in health state Λ_τ^1, the insurer's first period costs are the same as if a type θ consumer in health state Λ_τ^1 had chosen that contract, and the transitions to health states in the next period are the same as well. If the lapsation behavior starting in period $\tau+1$ were the same, the insurer would break even. But, we have just concluded that if the lapsation behavior is different, the insurer's expected continuation payoff must be weakly higher. Thus, the insurer must have a non-negative expected payoff when a type θ' consumer in health state Λ_τ^1 chooses contract $p_\tau^{\theta *}(\Lambda_\tau^1)$ in period τ. ■

We now prove Proposition 2:

Proof. of Proposition 2: We suppose that, in each period $t = 1, ..., T$, the menu of optimal guaranteed premium path contracts $\{p_t^{\theta *}(\Lambda_t^1)\}_{\theta \in \Theta, \Lambda_t^1 \in H_t}$ is offered, where $p_t^{\theta *}(\Lambda_t^1) \equiv \{y_t - c_t^{\theta *}(\Lambda_t^1)\}_{t=1}^T$. The proof is by induction. Consider the following induction hypothesis.

Induction Hypothesis: In each period $t > \tau$ the menu is self-selective and induces no secret savings: that is, if a consumer of type θ agrees to a new contract he chooses that type's optimal contract $p_t^\theta(\lambda_t)$ and engages in no secret savings.

The hypothesis is clearly true for $\tau = T-1$, as given any previous history the menu $\{p_T^\theta(\Lambda_T^1)\}$ is a singleton with $p_T = \mathbb{E}[m_T|\lambda_T]$, and hence necessarily self-selective, while there is no possibility of secret savings as period T is the last period. Now suppose it is true for some τ; we argue that it is then also true for $\tau-1$. Lemma 5 implies that a type θ consumer in health state $\Lambda_{\tau-1}^1$ choosing $p_{\tau-1}^{\theta *}(\Lambda_{\tau-1}^1)$ in period $\tau-1$ will not secretly save. From Lemma 6 we know that if a type θ consumer chooses in period $\tau-1$ when in health state $\Lambda_{\tau-1}^1$ the contract intended for him then insurers break even, but if he chooses instead the contract intended for type θ' then insurers earn non-negative profits (all future insurers break even in both cases). But the policy intended for the type θ consumer maximizes the type θ consumer's discounted expected utility subject to the constraint that insurers at least break-even (and the constraint that continuation profits can never be strictly positive). The policy intended for type θ' was therefore feasible for type θ, which implies that it cannot be preferred by type θ.

Applying induction, the menu $\{p_t^{\theta *}(\Lambda_1^1)\}_{\theta \in \Theta, \Lambda_1^1 \in H_1}$ is self-selective and induces no secret savings. ■

3 Appendix C: Proof of Proposition 3 (Consumer Inertia and Myopia)

The proof of Proposition 3 follows closely the proof of Proposition 1. Recall that we model inertia as a cost $\sigma > 0$ incurred by the consumer upon switching firms, which is equivalent to supposing that any new lapsation-inducing contract starts with subsidy $-\sigma$. To model myopia, we suppose that the consumer applies a discount factor $\beta < \gamma$ to future consumption, where γ is the discount factor of the insurers (and planner, when we conduct welfare analysis).

The proof strategy is based on induction: The result holds vacuously for $t = T$. Then, for $t < T$, we assume the proposition is true for the optimal contracts $c^*_{\Lambda^1_{t'}}(\cdot)$ at all $\Lambda^1_{t'}$ with $t' > t$, and show it is also true for the period-t optimal contracts $c^*_{\Lambda^1_t}(\cdot)$ for any Λ^1_t. To establish the result, we show that if for some Λ^1_t, the optimal contract $c^*_{\Lambda^1_t}(\cdot)$ does not satisfy (7), then there is a modification of $c^*_{\Lambda^1_t}(\cdot)$ that (i) is strictly preferred to $c^*_{\Lambda^1_t}(\cdot)$ by the consumer; and (ii) satisfies no-lapsation and zero-profit.

Before we get to the proof itself, we introduce notation on how to modify a contract.

Definition 5 *Let* $\min\{t', t''\} \geq t$. *We say contract* $\hat{c}_{\Lambda^1_t}(\cdot)$ *is an* ε*-transfer, from history* $\Lambda^1_{t'}$ *to history* $\Lambda^1_{t''}$*, on contract* $c_{\Lambda^1_t}(\cdot)$*, and write* $\hat{c}_{\Lambda^1_t}(\cdot) = tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^1_{t'}, \Lambda^1_{t''}]$ *if:*

1. $\hat{c}_{\Lambda^1_t}(\Lambda^1_{t'}) = c_{\Lambda^1_t}(\Lambda^1_{t'}) - \varepsilon$
2. $\hat{c}_{\Lambda^1_t}(\Lambda^1_{t''}) = c_{\Lambda^1_t}(\Lambda^1_{t''}) + \left[\varepsilon \times \frac{f(\Lambda^{t+1}_{t'}|\Lambda^1_t)}{f(\Lambda^{t+1}_{t''}|\Lambda^1_t)} \times \delta^{t'-t''}\right]$
3. *For all* $\tau \geq t$ *and* $\Lambda^t_\tau \notin \{\Lambda^t_{t'}, \Lambda^t_{t''}\}$*, we have* $\hat{c}_{\Lambda^1_t}(\Lambda^t_\tau) = c_{\Lambda^1_t}(\Lambda^t_\tau)$

In words, this ε-transfer just transfers some consumption between health histories $\Lambda^1_{t'}$ and $\Lambda^1_{t''}$ after applying a multiplier to the transfer to keep the discounted expected consumption, and hence insurer cost, unchanged. Our improvements on $c^*_{\Lambda^1_t}(\cdot)$ will be constructed using ε-transfers. We record two facts about such transfers:

Remark 3 ε*-transfers preserve the expected discounted profit: If* $c^*_{\Lambda^1_t}(\cdot) \in B^S(\Lambda^1_t)$ *for some* $S \in \mathbb{R}$*, then* $tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^1_{t'}, \Lambda^1_{t''}] \in B^S(\Lambda^1_t)$.

Remark 4 *For every* $\Lambda^t_{t'}$ *and* $\Lambda^t_{t''}$ *with* $\psi_{t'}\Big(c_{\Lambda^1_t}(\Lambda^1_{t'})\Big) > \psi_{t''}\Big(c_{\Lambda^1_t}(\Lambda^1_{t''})\Big)$ *there exists an* $\varepsilon_0 > 0$ *such that for all* $\varepsilon \leq \varepsilon_0$ *we have* $tr[c_{\Lambda^1_t}(\cdot), \varepsilon, \Lambda^t_{t'}, \Lambda^t_{t''}] \succ c_{\Lambda^1_t}(\cdot)$.

Remark 3 follows immediately from the fact that, using the insurers' discount factor δ, the ε-transfer does not change the expected discounted consumption in the contract, while Remark 4 follows because the consumer is strictly risk averse [$u(\cdot)$ is strictly concave].

Before proceeding to the proof of the proposition, we observe that in any optimal contract, the continuation contract specified at every future health history must itself be an optimal contract starting at that history for some subsidy:

Claim 2 *For $t' > t$, define $S_{t'}$ as the expected loss sustained by the insurer under contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ after the realization of health history $\Lambda^1_{t'} = \left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle$. Formally:*

$$S_{t'} = \Sigma^T_{\tau=t'} \delta^{\tau-t'} \left(\mathbb{E}[c^*_{\Lambda^1_t}(\left\langle \Lambda^1_t, \Lambda^{t+1}_\tau \right\rangle |S_t) - y_\tau - m_\tau |\Lambda^1_{t'}] \right).$$

Then, the following is true:

$$c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t) = c^*_{\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle}(\cdot|S_{t'}), \tag{29}$$

In words, Claim 2 states that any continuation contract $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t)$ of $c^*_{\Lambda^1_t}(\cdot|S_t)$ is in fact the optimal solution to the generalized problem outlined in Definition 1 for history $\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle$ when the subsidy available to the consumer is exactly the amount $S_{t'}$.

Proof of Claim 2. If at any continuation history $\Lambda^t_{t'}$ the condition in the claim did not hold we could replace the continuation contract $c^*_{\Lambda^1_t|\Lambda^{t+1}_{t'}}(\cdot|S_t)$ by $c^*_{\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle}(\cdot|S_{t'})$ and do strictly better for the consumer without violating no-lapsation or changing the required subsidy S_t for contract $c^*_{\Lambda^1_t}(\cdot|S_t)$, a contradiction to the optimality of $c^*_{\Lambda^1_t}(\cdot|S_t)$. ■

We now turn to proving Proposition 3. To do so, we will actually prove a more general statement than the proposition, using induction on the number of periods:

Lemma 7 *Consider optimal contract $c^*_{\Lambda^1_t}(\cdot|S_t)$. There exists a unique $\bar{c} \in \mathbb{R}$ such that $c^*_{\Lambda^1_t}(\Lambda^1_t|S_t) = \bar{c}$, and for any $t' > t$ and $\Lambda^{t+1}_{t'}$ such that $f(\Lambda^{t+1}_{t'}|\Lambda^1_t) > 0$, we have*

$$\psi_{t'}\left(c^*_{\Lambda^1_t}(\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle |S_t)\right) = \max\{\psi_t(\bar{c}), \psi_{t'}\left(c^*_{\left\langle \Lambda^1_t, \lambda_{t+1} \right\rangle}(\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle | -\sigma)\right)\}. \tag{30}$$

In words, Lemma 7 says that at any subsequent period t' and history $\left\langle \Lambda^1_t, \Lambda^{t+1}_{t'} \right\rangle$, contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ gives –after applying the myopic transformation– the larger value between (i) consumption that it immediately gives, and (ii) the consumption that the optimal, break-even contract with subsidy $-\sigma$ signed in the beginning of period $t+1$ at history $\langle \Lambda^1_t, \lambda_{t+1} \rangle$ would offer.

Remark 5 *Note that condition (30) of Lemma 7 implies that any two optimal contracts signed at time t and health history Λ^1_t, but with differing subsidies $S''_t > S'_t$, are ordered* *by the dominance relation according to the level of the initial consumptions they specify, which by the break-even condition are ordered according to the size of the subsidies; that is,* $c^*_{\Lambda^1_t}(\cdot|S''_t) \geq c^*_{\Lambda^1_t}(\cdot|S'_t)$*, with strict inequality at the initial history* Λ^1_t.

Proof of Lemma 7. The proof goes by induction. For $t = T$ the result is immediate: given that there is no period after $t = T$, condition (30) holds vacuously, and at that point $c_{\Lambda^1_T}(\Lambda^1_T|S) = y_T - \mathbb{E}[m_T|\lambda_T] + S$. We now turn to the proof for $t < T$, assuming, by way of induction, that the result holds for any $\tau > t$ and any S_τ. We begin by showing that (30) holds for any period $t+1$ history $\langle \Lambda^1_t, \lambda_{t+1} \rangle$; i.e., that

$$\psi_{t+1}\Big(c^*_{\Lambda^1_t}(\big\langle \Lambda^1_t, \lambda_{t+1} \big\rangle |S_t)\Big) = \max\{\psi_t\Big(c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)\Big), \psi_{t+1}\Big(c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\big\langle \Lambda^1_t, \lambda_{t+1} \big\rangle \,|-\sigma)\Big)\}. \quad (31)$$

To this end, we consider two cases regarding history nodes ending at period $t+1$.

Case 1. *At history* $\langle \Lambda^1_t, \lambda_{t+1} \rangle$*, the no-lapsation condition is binding for contract* $c^*_{\Lambda^1_t}(\cdot|S_t)$. Formally:

$$V_{\Lambda^1_t}(c^*_{\Lambda^1_t|\lambda_{t+1}}(\cdot|S_t)) = V_{\Lambda^1_t}(c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\cdot|-\sigma)). \quad (32)$$

Note that by Claim 2, the continuation contract $c^*_{\Lambda^1_t|\lambda_{t+1}}(\cdot|S_t)$ is itself the optimal contract $c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\cdot|S_{t+1})$ for some S_{t+1}. By Remark 5, (32) implies that this $S_{t+1} = -\sigma$; i.e., that

$$c^*_{\Lambda^1_t|\lambda_{t+1}}(\cdot|S_t) = c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\cdot|-\sigma), \quad (33)$$

which implies that

$$\psi_{t+1}\Big(c^*_{\Lambda^1_t}(\big\langle \Lambda^1_t, \lambda_{t+1} \big\rangle |S_t)\Big) = \psi_{t+1}\Big(c^*_{\langle \Lambda^1_t, \lambda_{t+1} \rangle}(\big\langle \Lambda^1_t, \lambda_{t+1} \big\rangle \,|-\sigma)\Big). \quad (34)$$

Next, note that the immediate consumption $c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)$ in contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ must satisfy

$$\psi_t\Big(c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)\Big) \leq \psi_{t+1}\Big(c^*_{\Lambda^1_t}(\big\langle \Lambda^1_t, \lambda_{t+1} \big\rangle |S_t)\Big), \quad (35)$$

for otherwise an ε-transfer from the immediate consumption at history Λ^1_t to history $\langle \Lambda^1_t, \lambda_{t+1} \rangle$ would strictly improve the contract $c^*_{\Lambda^1_t}(\cdot|S_t)$ from the perspective of the consumer, without changing the expected profit. This transfer would also satisfy no-lapsation, given that it weakly increases the consumption given by $c^*_{\Lambda^1_t}(\cdot|S_t)$ at any history that happens strictly after time t. Thus, (31) holds at all period-$(t+1)$ histories at which lapsation binds.

Case 2. *At history* $\langle \Lambda^1_t, \lambda_{t+1} \rangle$*, the no-lapsation condition is not binding for contract* $c^*_{\Lambda^1_t}(\cdot|S_t)$. That is,

$$V_{\Lambda_t^1}(c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t)) > V_{\Lambda_t^1}(c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\cdot|-\sigma)). \tag{36}$$

As in the previous case, given Claim 2, the continuation contract $c^*_{\Lambda_t^1|\lambda_{t+1}}(\cdot|S_t)$ is itself the optimal contract $c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\cdot|S_{t+1})$ for some S_{t+1}. By Remark 5, inequality (36) implies that:

$$c^*_{\Lambda_t^1|\lambda_{t+1}}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t) > c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|-\sigma). \tag{37}$$

or, equivalently, that

$$\psi_{t+1}\Big(c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t)\Big) > \psi_{t+1}\Big(c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|-\sigma)\Big) \tag{38}$$

We now claim that

$$\psi_t\Big(c^*_{\Lambda_t^1}(\Lambda_t^1|S_t)\Big) = \psi_{t+1}\Big(c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t)\Big). \tag{39}$$

To see this, note that if $\psi_t\Big(c^*_{\Lambda_t^1}(\Lambda_t^1|S_t)\Big) < \psi_{t+1}(c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t))$, then an ε-transfer from the history $(\Lambda_t^1,\lambda_{t+1})$ to the immediate history Λ_t^1 will increase the consumer's expected utility, will not change the expected profit from the contracts, and will preserve no-lapsation if ε is small enough, given (36). This contradicts the assumption that $c^*_{\Lambda_t^1}(\cdot|S_t)$ is the optimal contract. Conversely, if $\psi_t\Big(c^*_{\Lambda_t^1}(\Lambda_t^1|S_t)\Big) > \psi_{t+1}(c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t))$, the reverse ε-transfer will strictly increase the consumer's expected utility and preserve the insurer's expected profit. It also preserves no-lapsation since it weakly increases consumption at any history strictly after Λ_t^1. Conditions (38) and (39) imply that (31) also holds at all period-$(t+1)$ histories at which lapsation does not bind.

To sum up, cases 1 and 2 show that for *any* history λ_{t+1}, no matter whether no-lapsation is binding or not, (31) holds. Next, we combine (31) with the induction assumption to extend the argument, which currently applies only to period-$(t+1)$ histories $\langle\Lambda_t^1,\lambda_{t+1}\rangle$, also to any history $\langle\Lambda_t^1,\Lambda_{t'}^{t+1}\rangle$ with $t' > t+1$. By Claim 2, we know that for some appropriate S_{t+1}, and any $(\Lambda_t^1,\lambda_{t+1},\Lambda_{t'}^{t+2})$, we have:

$$c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_t) = c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\langle\Lambda_t^1,\lambda_{t+1}\rangle|S_{t+1}) \tag{40}$$

and

$$c^*_{\Lambda_t^1}(\langle\Lambda_t^1,\lambda_{t+1},\Lambda_{t'}^{t+2}\rangle|S_t) = c^*_{\langle\Lambda_t^1,\lambda_{t+1}\rangle}(\langle\Lambda_t^1,\lambda_{t+1},\Lambda_{t'}^{t+2}\rangle|S_{t+1}) \tag{41}$$

By induction, we know that

$$\psi_{t'}\Big(c^*_{\langle\Lambda^1_t,\lambda_{t+1}\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|S_{t+1})\Big) =$$

$$\max\{\psi_{t+1}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle\,|S_{t+1})\Big),\psi_{t'}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1},\lambda_{t+2}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|-\sigma)\Big)\} \quad (42)$$

Replacing into equation (42) from (40) and (41), we get:

$$\psi_{t'}\Big(c^*_{\Lambda^1_t}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|S_t)\Big) =$$

$$\max\{\psi_{t+1}\Big(c^*_{\Lambda^1_t}(\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle\,|S_t)\Big),\psi_{t'}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1},\lambda_{t+2}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|-\sigma)\Big)\} \quad (43)$$

Now, substituting for $c^*_{\Lambda^1_t}(\langle\Lambda^1_t,\lambda_{t+1}\rangle\,|S_t)$ from (31), we get:

$$\psi_{t'}\Big(c^*_{\Lambda^1_t}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|S_t)\Big) \quad (44)$$

$$= \max\{\max\{\psi_t\Big(c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)\Big),\psi_{t+1}\Big(c^*_{(\Lambda^1_t,\lambda_{t+1})}(\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle\,|-\sigma)\Big)\},\psi_{t'}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1},\lambda_{t+2}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|-\sigma)\Big)\}$$

$$= \max\{\psi_t\Big(c^*_{\Lambda^1_t}(\Lambda^1_t|S_t)\Big),\max\{\psi_{t+1}\Big(c^*_{(\Lambda^1_t,\lambda_{t+1})}(\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle\,|-\sigma)\Big),\psi_{t'}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1},\lambda_{t+2}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|-\sigma)\Big)\}\}$$

But by our induction assumption, the inner maximum equals $\psi_{t'}\Big(c^*_{\big\langle\Lambda^1_t,\lambda_{t+1}\big\rangle}(\big\langle\Lambda^1_t,\lambda_{t+1},\Lambda^{t+2}_{t'}\big\rangle\,|-\sigma)\Big)$; substituting this into (44) tells us that (30) holds for contracts signed in period t. Applying induction establishes the lemma.■

Applying Lemma 7 to the special case of $S_t = 0$, we get that for any Λ^1_t and $\Lambda^{t+1}_{t'}$ such that $f(\Lambda^{t+1}_{t'}|\Lambda^1_t) > 0$, we have

$$\psi_{t'}\Big(c^*_{\Lambda^1_t}(\big\langle\Lambda^1_t,\Lambda^{t+1}_{t'}\big\rangle)\Big) = \max\{\psi_t\Big(c^*_{\Lambda^1_t}(\Lambda^1_t)\Big),\psi_{t'}\Big(c^*_{\Lambda^1_{t+1}}(\big\langle\Lambda^1_t,\Lambda^{t+1}_{t'}\big\rangle\,|-\sigma)\Big)\}$$

Since Lemma 7 holds for $c^*_{\Lambda^1_{t+1}}(\cdot|-\sigma)$ as well, we can expand $c^*_{\Lambda^1_{t+1}}(\big\langle\Lambda^1_t,\Lambda^{t+1}_{t'}\big\rangle\,|-\sigma)$ in the same way as above, substituting into (3

$$\psi_{t'}\Big(c^*_{\Lambda^1_{t+1}}(\big\langle\Lambda^1_t,\Lambda^{t+1}_{t'}\big\rangle\,|-\sigma)\Big) = \max\{\psi_{t+1}(c^*_{\Lambda^1_{t+1}}(\Lambda^1_{t+1}|-\sigma)),\psi_{t'}\Big(c^*_{\Lambda^1_{t+2}}(\big\langle\Lambda^1_t,\Lambda^{t+1}_{t'}\big\rangle\,|-\sigma)\Big)\}. \quad (45)$$

Then we can do this again and again, until we get that for all $t' > t$ and $\Lambda^t_{t'}$ such that $f(\Lambda^t_{t'}|\Lambda^1_t) > 0$,

$$\psi_{t'}\left(c^*_{\Lambda^1_t}(\Lambda^1_{t'})\right) = \max\{\psi_{t'}\left(c^*_{\Lambda^1_t}(\Lambda^1_t)\right), \max_{\tau\in\{t+1,\dots,t'\}} \psi_\tau\left(c^*_{\Lambda^1_\tau}(\Lambda^1_\tau)| - \sigma\right)\}, \tag{46}$$

which is exactly the statement made in the proposition.■

4 Appendix D: PKH premiums and Cochrane contracts

In this appendix we discuss further PKH guaranteed renewable contracts, and also Cochrane (1995)'s premuium insurance scheme. We end by discussing the empirical difference in initial premia and welfare between these contracts and our optimal contracts.

To begin, we first derive a general formula (for arbitrary T) for the premia in a PKH guaranteed renewable contract. We show that, in the context of our model, these policies provide a consumer who starts at age 25 in the healthiest possible state with the guaranteed consumption path $\{y_t - p_t\}^T_{t=1}$, where the period t premium p_t is (in anticipation of our empirical analysis where the health process is second-order Markov, we denote by $\Lambda_t = (\lambda_{t-1}, \lambda_t) = (1, 1)$ the healthiest possible state in period t):[10]

$$p_t = \mathbb{E}[m_t|\Lambda_t = (1,1)] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau|\Lambda_t = (1,1)] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \text{ for } t = 1, ..., T \tag{47}$$

To this end, consider one-period contracts signed in each period t in return for the premium $p_t(\Lambda_t)$ paid at signing that does the following:

- fully insures period t health expenses
- if $t < T$, pays in addition the amount $p_{t+1}(\Lambda_{t+1}) - p_{t+1}(1, 1)$ [where $p_{t+1}(1, 1)$ is the period $t + 1$ premium for the healthiest period $t + 1$ health state, $\Lambda_{t+1} = (1, 1)$, at the start of the next period $t + 1$].

These contracts pay an amount that guarantees that the insured's outlays for the next period contract (net of the insurance payout from the previous period) always equal the amount that the healthiest type would pay.

The premiums for these contracts will in equilibrium be:

$$p_T(\Lambda_T) = \mathbb{E}[m_T|\Lambda_T]$$

[10]PKH focus on the case in which the consumer starts in the healthiest possible state.

and for $t < T$,

$$p_t(\Lambda_t) = \mathbb{E}[m_t|\Lambda_t] + \delta\mathbb{E}\{p_{t+1}(\Lambda_{t+1}) - p_{t+1}(1,1)|\Lambda_t\}$$

Lemma 8 *For all t, $p_t(\Lambda_t) = \mathbb{E}[m_t|\Lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_t] - \mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+1} = (1,1)]\}$*

Proof. Clearly true in period T. Suppose it is true for all periods $\tau > t$. To see it is true in period t, we substitute and use the Law of Iterated Expectations:

$$\begin{aligned}
p_t(\Lambda_t) &= \mathbb{E}[m_t|\Lambda_t] + \delta\{\mathbb{E}[p_{t+1}(\Lambda_{t+1})|\Lambda_t] - p_{t+1}(1,1)\} \\
&= \mathbb{E}[m_t|\Lambda_t] + \delta\mathbb{E}\{\mathbb{E}[m_{t+1}|\Lambda_{t+1}] + \sum_{\tau>t+1} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+1}] - \mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+2} = (1,1)]|\Lambda_t\} \\
&\quad -\delta\{\mathbb{E}[m_{t+1}|\Lambda_{t+1} = (1,1)] + \sum_{\tau>t+1} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+1} = (1,1)] - \mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+2} = (1,1)]\} \\
&= \mathbb{E}[m_t|\Lambda_t] + \delta\{\mathbb{E}[m_{t+1}|\Lambda_t] + \sum_{\tau>t+1} \delta^{\tau-(t+1)}\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_t] \\
&\quad -\delta\{\mathbb{E}[m_{t+1}|\Lambda_{t+1} = (1,1)] + \sum_{\tau>t+1} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+1} = (1,1)]\} \\
&= \mathbb{E}[m_t|\Lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_t] - \mathbb{E}[m_\tau(\Lambda_\tau)|\Lambda_{t+1} = (1,1)]\}
\end{aligned}$$

■

Cochrane (1995) proposes a different scheme to protect consumers from reclassification risk: premium insurance purchased in each period t that pays the consumer the change in the present discounted value of his future medical expenses at the start of the following period, equal to

$$\sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_{t+1}] - \mathbb{E}[m_\tau|\Lambda_t]\},$$

which can potentially yield first-best insurance. In principle, in this manner, first-best insurance could be provided to the consumer. As Cochrane notes, however, this policy has the problem that the consumer would have to pay the insurer when the evolution of his expected future health expenses is better than expected, which may be impossible to enforce. Cochrane (1995) proposes to solve this problem via health savings accounts that can be used to receive and make these premium insurance payments. Unfortunately, such an account can hit a zero balance because a consumer who starts healthy ($\Lambda_{25} = (1,1)$) and remains healthy ($\Lambda_t = (1,1)$ for all $t > 1$) would need to make payments in every period. (That is, remaining healthy is a *better than expected* outcome that requires the consumer to pay the insurance company.)

An alternative approach that one might consider to avoid the need for consumer end-of-period repayments in the premium insurance scheme would have the consumer pre-pay the maximal possible repayment at the start of the period as part of his premium. That is, in each period t, the consumer would pay a total premium, including for both medical insurance and premium insurance, equal to

$$\mathbb{E}[m_t|\Lambda_t] + \sum_{\tau>t} \delta^{\tau-t}\{\mathbb{E}[m_\tau|\Lambda_t] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \tag{48}$$

and, in addition to coverage of period t medical claims, in each period $t+1$ (for $t < T$) the insurer would pay the consumer the non-negative amount

$$\begin{aligned} \text{Payment} &= \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_{t+1}] - \mathbb{E}[m_\tau|\Lambda_t]\} + \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_t] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \\ &= \sum_{\tau>t} \delta^{\tau-(t+1)}\{\mathbb{E}[m_\tau|\Lambda_{t+1}] - \mathbb{E}[m_\tau|\Lambda_{t+1} = (1,1)]\} \geq 0, \end{aligned} \tag{49}$$

equal to the change in expected medical expenses plus the repayment (with interest) of the second term in (48). Subtracting the period t payment [given by expression (49) modified to be for period t rather than $t+1$] from the period t premium (48), we see that the net premium payment in each period t for a consumer who begins with $\Lambda_{25} = (1,1)$ is exactly the PKH premium (47). Thus, this approach to premium insurance is exactly equivalent to a PKH guaranteed renewable contract, and hence would give the insured lower discounted expected utility than our optimal dynamic contract.

4.1 Empirical Comparison of PKH and Optimal Dynamic Contracts

Using formula (47), we calculate that for our Utah male sample the initial PKH premium paid by a healthy 25 year old [i.e., a consumer with $\Lambda_{25} = (1,1)$] is about 3.2% higher than the initial premium paid by a healthy 25 year old individual with flat net income in the optimal dynamic contract.[11] For a consumer who arrives at age 25 in the healthiest state and who has a flat net income profile, the excessively low initial consumption required to eliminate all reclassification risk translates into a lower welfare: $CE_{PKH} = \$54,834$, which is 0.4% lower than the certainty equivalent that this consumer would have with an optimal dynamic contract. As a result, the PKH contract eliminates 97.2% of the welfare loss from reclassification, compared to the 99.4% from an optimal dynamic contract. The welfare loss from the PKH contract relative to an optimal contract increases with rising income profiles:

[11] At age 25 the value of the second term in equation (47), representing the premium pre-payment that is required in the PKH contract, is \$1,530. This amount divided by δ $(= 0.975)$ is also the end-of-period amount that the consumer would need to pay out in the event that she remained healthy (with $\Lambda_{26} = (1,1)$) to achieve the first best in the reclassification-risk insurance scheme proposed by Cochrane (1995).

For example, for a healthy 25 year old downscaled manager we find that $CE_{PKH} = \$37,819$, resulting in a loss of 2.4% compared to an optimal dynamic contract; the PKH contract therefore eliminates only 84.2% of the welfare loss due to reclassification risk, compared to the 94.3% from an optimal dynamic contract. For a non-manager $CE_{PKH} = \$47,525$, which represents a 1.5% welfare loss and an elimination of 80.1% of the welfare loss from reclassifcation risk, compared to 95.1% from an optimal contract.[12]

5 Appendix E: Extending the model to capture partial access to credit markets

Our original submission focused on the benchmark with no external borrowing (outside of the dynamic insurance contract). This appendix adds a framework that allows partial access to credit markets to allow for the realistic scenario where consumers can have some limited independent borrowing. This allows them to smooth income over time under increasing income paths and, in turn, may unlock the benefits of dynamic contracts by allowing for increased front-loading and, thus, increased insurance of reclassification risk.

5.1 Theory

We extend the model in the following way to capture the possibility of borrowing: In the beginning of each period history Λ_t^1, the customer makes two simultaneous decisions. First, whether to stay with her current insurance contracts or lapse to a new one offered by the market; second, how much to borrow. To formalize the borrowing decisions, we provide two more definitions.

Definition 6 *We denote a "borrowing portfolio" by a function $b(\cdot)$, from the set of all possible pairs histories Λ_t^1 and future periods $t' > t$ to the set of non-negative numbers. Therefore, $b(\Lambda_t^1, t')$ is the amount the individual decides to borrow in the beginning of health history Λ_t^1, which she returns with interest to the lending institution at the beginning of period $t' > t$. That is, she returns $\frac{b(\Lambda_t^1, t')}{\delta^{t'-t}}$*

Definition 7 *The "Maximum available borrowing portfolio" is denoted by $\bar{b}(\cdot)$. For each (Λ_t^1, t') with $t' > t$, the value $\bar{b}(\Lambda_t^1, t')$ is the highest possible value for $b(\Lambda_t^1, t')$.*

This latter definition allows us to capture the idea that access to credit markets is "limited."

[12] While the PKH contract assumes that the consumer arrives at age 25 in the healthiest state, as we have seen, not all consumers manage to do so. We expect that the excessive front-loading involved in contracts that would eliminate al reclassification risk would be more costly for such consumers.

The main question here is whether the simultaneity and interplay between the "insurance problem" and the "borrowing problem" from the perspective of the consumer leads to complications that would lead our optimal contract results to not hold anymore. The result below says the answer is no.

Proposition 5 *There is at least one optimal solution to the dual insurance-borrowing dynamic problem in which the consumer solves the two problems separately. Specifically:*

1. *She borrows according to portfolio $b^*(\cdot) = \bar{b}(\cdot)$. That is she borrows as much as she can.*

2. *Then she signs up for a different optimal contract $c^{\theta,\bar{b}*}_{\Lambda^1_1}(\cdot)$ based on her borrowing portfolio, rather than $c^{\theta*}_{\Lambda^1_1}(\cdot)$.*

Multiple optimal solutions may exist but they all lead to the same consumption in each history Λ^1_t that happens with positive probability.

Remark 6 *In the special case where $\bar{b}(\Lambda^1_t, t')$ can only depend on t and t' (that is, when borrowing restrictions are health-independent), the optimal borrowing strategy by the consumer simply leads to a different income profile $\bar{y}$. Therefore, in the second step in Proposition 5, we can write $c^{\theta,\bar{b}*}_{\Lambda^1_1}(\cdot) = c^{\bar{\theta}*}_{\Lambda^1_1}(\cdot)$ where $\bar{\theta} = (\bar{y}, u)$. This means from a computational perspective, we have two steps to compute the equilibrium: first compute $\bar{y}$, second use our existing algorithm to compute the equilibrium for the new state $\bar{\theta}$*

Proof Idea for Proposition 5. We skip a formal proof of this proposition but it will be available upon request. On an intuitive level, the customer borrows as much as she can because she can always "send the borrowed money back to the future through the dynamic contract" by front-loading all of the borrowed money. Doing so can never harm the customer's welfare and is in fact likely to be strictly preferable to not borrowing. This is because when borrowing, the customer transfers money from all health states of her t' self to time t. But when front-loading the money and sending it back to period t' as a consumption guarantee, the healthier t' selves of the customer receive less than what was borrowed from them whereas the less healthy ones receive more (by the definition of consumption guarantees). As such, the customer will optimally combine borrowing and long term insurance in order to smooth out her period t' consumption across health states (note that the customer does not always front-load all of the borrowed amount, especially if her income profile is steep).

Borrowing and First-Year Equilibrium Contract Terms: Manager Income							
				γ			
	0	0.1	0.2	0.3	0.4	0.5	0.6
Premium	1.072	1.340	1.435	1.496	1.524	1.562	1.591
First-Year Costs	0.837	0.837	0.837	0.837	0.837	0.837	0.837
Front-Loading	0.235	0.503	0.597	0.659	0.687	0.725	0.754
Consumption	49.942	53.644	55.802	57.539	58.928	60.309	61.518

Table 11: First-year contract terms in the equilibrium long-run contract for men with a manager income path, showing first-year premiums, expected costs, the extent of front-loading, and consumption levels (thousands of dollars). This table is for health status $\lambda_{24} = \lambda_{25} = 1$ and covers range of credit market access parameter γ from 0 to 0.6.

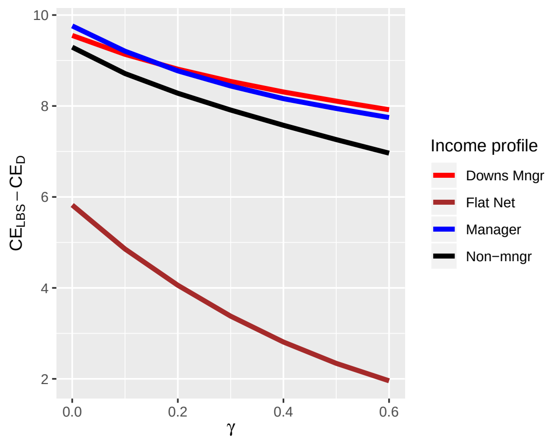
5.2 Empirical Analysis.

Although our theoretical analysis allows for a much more general model of credit availability, in our empirical analysis we focus on a more specific model. We assume that "access to credit markets" is governed by only one parameter γ. Consider an individual with income profile y. We assume that at the beginning of each year t, the individual can take out a loan as much as $\gamma \times y_t$, which is to be returned one year after (i.e., in the beginning of year $t+1$) with the interest rate $\frac{1}{\delta} - 1$. That is, the individual will have to return $\frac{\gamma y_t}{\delta}$ to the lender. As can be seen from this formulation γ indeed is a measure of access to credit markets. According to our theoretical result, the individual would always borrow the whole feasible amount of γy_t, effectively constructing a different income path $\bar{y}$. She then signs the contract that is optimal for her "new" type $(\theta, \bar{y})$.

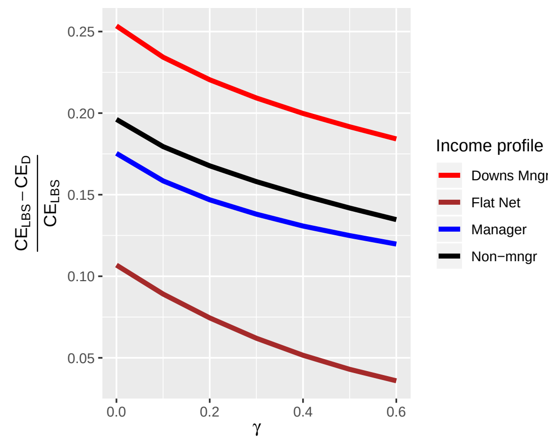
Table 11 shows what happens to the first year terms of the optimal contract as we move parameter γ. The better access the customer has to credit markets (i.e., the higher the γ), the more she consumes *and* frontloads in the first period. That is, she borrows more and uses some of the borrowed amount to purchase dynamic contracts that provide better insurance against reclassification risk.

While Table 11 described the response of the contract terms to γ, Figure 6 presents the welfare results. As this figure shows, the performance of one-sided commitment contracts, compared to two-sided commitment ones, improves as γ increases.[13] As γ increases and consumers can borrow more, dynamic contracts become closer in performance to the first-best

[13]Note that in figure 6, welfare under one sided commitment contracts CE_D is NOT compared to the no-saving-no-borrowing welfare CE_{NBNS}. It is, rather, compared to "limited saving and borrowing" CE_{LBS}. This latter measure allows the individual under two-sided commitment contracts to save as much as he would like to and combine it with the same borrowing scheme that the one-sided commitment customer is exposed to. This was done to ensure a fair comparison between one- and two-sided commitment systems.



(a) absolute



(b) relative

Figure 6: The gap between the performances of one-sided and two-sided commitment contracts lowers as access to credit market (measured by γ) increases. This is the case both in both absolute and relative terms as shown by panels a and b respectively.

contracts with two-sided commitment. The intuition is as follows: borrowing improves the welfare under two-sided commitment contracts only through aiding inter-temporal consumption smoothing. But it does so with one-sided commitment contracts through helping both with inter-temporal smoothing and with frontloading. These results give a quantitative sense of the extent to which borrowing can ease the costs of front-loading and enable more efficient contracting and improved insurance against reclassification risk with dynamic contracts with one-sided commitment.

6 Appendix F: Precautionary Savings

So far we have not allowed for savings in our welfare calculations. From Proposition 1 we know that this is without loss of generality for the case of optimal contracts with one-sided commitment. Consumers also would not want to engage in savings in the first best. However, with spot contracting consumers may want to engage in precautionary savings to lower the costs of reclassification risk. Individuals can save in good states to weather periods of bad health.

To study the impact of precautionary savings we solve a finite-horizon savings problem, with the same underlying fundamentals as in our main analysis, namely, the same income profiles, risk preferences, and transition matrices. We find optimal savings starting at age 25 given an income profile and the actuarially fair health insurance premiums associated

Income profile	(1) C^*_{NBNS}	(2) CE_{SPOT}	(3) CE_{SPOTwS}	(4) CE_D	(5) $\frac{C^*_{NBNS}-CE_{SPOTwS}}{C^*_{NBNS}}$	(6) $\frac{CE_{SPOTwS}-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$	(7) $\frac{CE_D-CE_{SPOTwS}}{C^*_{NBNS}-CE_{SPOTwS}}$
Flat net	54.50	44.18	46.98	48.66	0.138	0.271	0.223
Non-mngr	47.37	36.96	37.64	38.08	0.205	0.065	0.046
Manager	55.67	45.44	45.71	45.91	0.179	0.026	0.020
Downs Mngr	37.68	27.35	27.83	28.13	0.262	0.046	0.031

Table 12: Long-run welfare of Utah men allowing for precautionary savings under spot contracts, with a constant absolute risk aversion coefficient of 0.0004. Welfare measures in columns (1)–(4) are reported in thousands of dollars. The certainty equivalent of spot contracting with precautionary savings is denoted by CE_{SPOTwS}.

with the different health states.[14] Once we find optimal savings for each age and state, we compute the certainty equivalent, which we denote by CE_{SPOTwS} ($SPOTwS$ = "Spot with Savings").

Table 12 shows the welfare effect of precautionary savings in the Utah male sample. As the spot contracting with precautionary savings outcome is feasible in our dynamic problem with one-sided commitment, CE_{SPOTwS} naturally lies between CE_{SPOT} and CE_D. Savings enable the consumer to transfer resources to future periods, to be consumed in periods of high marginal utility from consumption. While these precautionary savings reduce the losses from reclassification risk, these losses remain very high, ranging between 13.8% and 26.2% of lifetime certainty equivalent (see column (5)). Optimal dynamic contracts do better than precautionary savings, as they allow for state-specific savings. By charging state-contingent premiums the optimal contract enables equating consumption across all states in which the lapsation does not bind.

Column (6) shows that precautionary savings closes a relatively small share – between 2.6% and 27.1% – of the welfare gap between spot contracts without savings and the no-borrowing/no-saving benchmark. Column (7) shows the fraction of the welfare gap between the no-borrowing/no-saving constrained first-best outcome and the spot contracting with precautionary savings outcome that is closed by optimal dynamic contracts; this ranges from 22.3% for flat net income profiles to 2.0% for managers.

[14]For each income profile, we solve a finite-horizon dynamic programming problem, from ages 25 to 65. Starting at age 64, for a grid of saving values entering that period, the individual finds the optimal saving level going into the last period that maximizes the sum of current utility from consumption and the discounted value of the expected utility in the last period, where the expectation is taken for each state given the transition matrices. Once we obtain the value function at age 64 for each possible health state and incoming saving level, we proceed backwards all the way to age 25, where we obtain the discounted expected utility starting in each possible health state. The ex-ante certainty equivalent is the certain consumption level that makes the consumer indifferent to the expected utility of entering the dynamic problem before observing the health realization at age 25.

λ_{t-1}	λ_t	λ_{t+1} 1	2	3	4	5	6	7
1	1	0.66	0.15	0.09	0.04	0.02	0.02	0.01
7	1	0.59	0.18	0.23	0	0	0	0
1	7	0.28	0.13	0.16	0.11	0.09	0.10	0.13
7	7	0.01	0.01	0.02	0.04	0.05	0.13	0.74

Health status (age 44 and 45)	Ages 45	46-50	51-55	56-64
1	0.84	2.27	4.07	5.35
4	3.05	4.51	5.19	5.70
7	20.51	13.05	7.90	6.38

Table 13: The top panel gives an example of empirical health status transitions from one year to the next, for 40-45 year old men. The bottom panel reports, for various age ranges, the constant annual medical expenses (in thousands of dollars) such that the present discounted value of these constant annual expenses equals the expected present discounted value of expenses over the age range in question for a Utah man in various age-40 health states.

7 Appendix G: Additional Tables and Figures

This appendix provides additional tables and figures to complement some of the analyses in the main text of the paper.

7.1 Health Transition Probabilities and Persistence

Table 3 in the main text depicted some health status transition probabilities for individuals between 30 and 35 years old. Table 13 provides the same information for those between ages of 40 and 45. As can be seen from comparing the two tables, *ceteris paribus*, the probability of transitioning to sicker future health states is higher for older individuals.

7.2 Second-Year Consumption Analysis for Flat Net Income Path

Table 14 shows period two consumption levels for consumers with flat net income paths under our baseline model. This table is the analog to the period two continent premium analysis presented in the text in Table 5.

Second-Year Equilibrium Consumptions: Flat Net Income								First-Year
			λ_{26}					Consumption
$\lambda_{24} = \lambda_{25}$	1	2	3	4	5	6	7	
1	54.791	54.765	54.765	54.765	54.765	54.765	54.765	54.765
4	54.724	54.13	52.739	51.189	50.803	50.803	50.803	50.803
7	54.809	54.724	54.544	52.219	41.384	42.396	36.683	36.548

Table 14: First- and second-year consumptions in the equilibrium long-run contract for men with a flat net income path, as a function of the period 1 health state and period 2 health status (thousands of dollars).

Second-Year Equilibrium Premiums: Downscaled Managers								First-Year
				λ_{26}				Premium
$\lambda_{24} = \lambda_{25}$	1	2	3	4	5	6	7	
1	1.160	1.873	2.605	3.012	3.012	3.012	3.012	1.165
4	0.884	1.405	2.548	3.490	5.334	5.334	5.334	3.487
7	0.843	1.375	1.973	3.054	11.895	11.851	20.511	20.511

Table 15: First- and second-year premiums in the equilibrium long-run contract for men with a downscaled manager income path, as a function of the period 1 health state and period 2 health status (thousands of dollars).

7.3 Second-year Premium and Consumption Analysis for the Downscaled Manager Income Path

Tables 15 and 16 show second-year (age-26) premiums and consumption levels for downscaled managers as a function of different health histories. Though front-loading is much more limited, the age-26 health states in which the lapsation constraint binds, conditional on the initial age-25 health state, are quite similar to those of consumers with flat net income profiles.

Second-Year Equilibrium Consumption: Downscaled Managers								First-Year
				λ_{26}				Consumption
$\lambda_{24} = \lambda_{25}$	1	2	3	4	5	6	7	
1	34.372	33.659	32.927	32.520	32.520	32.520	32.520	32.520
4	34.648	34.127	32.984	32.042	30.199	30.199	30.199	30.199
7	34.689	34.158	33.559	32.478	23.637	23.681	15.021	13.175

Table 16: First- and second-year consumptions in the equilibrium long-run contract for men with a downscaled manager income path, as a function of the period 1 health state and period 2 health status (thousands of dollars).

Init health	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\lambda_{24} = \lambda_{25}$	C^*	C^*_{NBNS}	CE_{SPOT}	CE_{SS}	CE_D	$\frac{C^*_{NBNS} - CE_{SPOT}}{C^*_{NBNS}}$	$\frac{CE_D - CE_{SPOT}}{C^*_{NBNS} - CE_{SPOT}}$	$\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$
1	55.14	48.68	41.39	43.78	48.26	0.150	0.943	0.500
2	54.96	48.11	39.78	42.74	47.41	0.173	0.915	0.502
3	54.84	47.64	39.31	42.62	46.84	0.175	0.903	0.485
4	54.36	46.77	40.23	42.39	45.42	0.140	0.794	0.367
5	52.86	43.63	34.91	36.77	40.41	0.200	0.631	0.307
6	51.51	40.21	33.42	35.33	37.69	0.169	0.629	0.236
7	49.33	30.71	29.78	29.85	29.88	0.030	0.105	0.005

Table 17: Long-run welfare results showing the certainty equivalent annual consumption of different insurance institutions under various initial health states, the non-manager income profile, a discount factor of 0.975, and constant absolute risk aversion equal to 0.0004. Units in columns (1)-(4) are 1000s of dollars.

Init health	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\lambda_{24} = \lambda_{25}$	C^*	C^*_{NBNS}	CE_{SPOT}	CE_{SS}	CE_D	$\frac{C^*_{NBNS} - CE_{SPOT}}{C^*_{NBNS}}$	$\frac{CE_D - CE_{SPOT}}{C^*_{NBNS} - CE_{SPOT}}$	$\frac{CE_D - CE_{SPOT}}{C^* - CE_{SPOT}}$
1	85.47	57.12	51.81	53.19	56.87	0.093	0.954	0.150
2	85.28	56.51	49.60	51.85	55.97	0.122	0.922	0.179
3	85.17	55.95	48.96	51.35	55.36	0.125	0.917	0.177
4	84.68	55.04	52.53	53.54	54.54	0.046	0.801	0.063
5	83.18	52.68	44.97	46.68	50.05	0.146	0.658	0.133
6	81.83	49.40	43.21	45.03	47.10	0.125	0.629	0.101
7	79.66	38.16	37.75	37.76	37.77	0.011	0.051	0.001

Table 18: Long-run welfare results showing the certainty equivalent annual consumption of different insurance institutions under various initial health states, the manager income profile, a discount factor of 0.975, and constant absolute risk aversion equal to 0.0004. Units in columns (1)-(4) are 1000s of dollars.

7.4 Additional Welfare Results Conditional on a Consumer's Age-25 Health State

7.5 First and Second-period Equilibrium Consumption Levels under Switching Costs

Table 19 shows first and second-year consumption levels for a flat net income profile and switching costs of $1,000 in the Utah male data. Comparing to Table 5, it is interesting to note that for all second-year states without a binding lapsation constraint consumption is higher with a higher switching cost, while consumption is lower for second-year states with a binding lapsation constraint. Namely, conditional on a history, higher switching costs enable transferring resources from the good to the bad states.

		Second-year consumption							
		λ_{26}							First-year
λ_{24}	λ_{25}	1	2	3	4	5	6	7	consumption
1	1	54.816	54.816	54.816	54.816	54.816	54.816	54.816	54.816
4	4	54.698	54.105	52.714	51.214	51.214	51.214	51.214	51.214
7	7	54.783	54.699	54.519	52.194	41.359	42.371	36.996	36.996

Table 19: First and second-year consumptions (in $1,000s) for Utah men with switching costs of $1,000, flat net income, and a constant absolute risk aversion coefficient equal to 0.0004.

8 Appendix H: Lapsation to Uninsurance

Our analysis in the main text of the paper considered only for lapsation into competing long-term health insurance contracts, as gains from trade would always exist from continuing insurance coverage. In reality, however, lapsation into uninsurance also takes place in the market (Konetzka and Luo, 2010). Here we discuss an extension of our model that takes this possibility into account. Lapsation to uninsurance may happen for reasons ranging from consumer inattention to the need to renew to negative income shocks that restrict the consumer's liquidity. In this appendix, however, we model such lapsations as an exogenous phenomenon most closely resembling lapsation due to inattention.

Formally, we examine an extension of our model in which at each period, there is a chance of γ that the consumer lapses to uninsurance. In case this lapsation takes place, the consumer will not return to the insurance market. We assume this lapsation chance is not a function of what period t the consumer is at or of her health status λ_t at that period. We analyze what the equilibrium contract will look like under this extension of our model. Proposition 6 below provides the answer.

Proposition 6 *If consumers lapse with an exogenous probability γ in each period, the optimal contract takes the same form as that described in Proposition 1 except that in the calculations, the common discount factor δ should be replaced with $\delta' \equiv \delta \times (1 - \gamma)$.*

Sketch of Proof for Proposition 6. We omit the details for the proof and confine ourselves to providing two remarks based on which the proof is built.

Remark 7 *Consider two contracts $c^1_{\Lambda^1_t}(\cdot)$ and $c^2_{\Lambda^1_t}(\cdot)$ offered to the consumer at the beginning of period t. If the consumer does not lapse from either (except for the exogenous lapsation into uninsurance), then the difference between the net present values of his expected utilities from the two contracts will be given by:*

$$\Sigma_{\tau=t}^{T}\hat{\delta}^{\tau-t}\mathbb{E}\big[u\big(c^{1}_{\Lambda^1_t}(\Lambda^t_\tau|\Lambda^1_t)\big)-u\big(c^{2}_{\Lambda^1_t}(\Lambda^t_\tau|\Lambda^1_t)\big)\big]$$

This remark simply describes how the consumer takes into account possible future lapsations into uninsurance as he compares different insurance contracts against one another. It follows because the consumer's continuation utility after an exogenous lapsation is the same regardless of the contract he is lapsing from. The remark states that in making such comparisons, the consumer behaves as if he will not be lapsing but instead has a discount factor of $\hat{\delta}=\delta\times(1-\gamma)$.

Remark 8 *Consider contract $c_{\Lambda^1_t}(\cdot)$ offered to the consumer at the beginning of period t. If the consumer does not lapse from this contract (except for the exogenous lapsation into uninsurance) then the insurer's expected profit from the contract is given by:*

$$\pi=\Sigma_{\tau=t}^{T}\hat{\delta}^{\tau-t}\mathbb{E}\big[y_\tau-m_\tau-c_{\Lambda^1_t}(\Lambda^t_\tau|\Lambda^1_t)\big]$$

This remark simply states that under an exogenous probability γ of lapsation into uninsurance, the insurer's expected profit from a contract is the same as what it would be if there was no lapsation to uninsurance but instead the insurer discounted future profits at the rate of $\hat{\delta}=\delta\times(1-\gamma)$. It follows because the insurer earns zero following any exogenous lapsation, regardless of the contract lapsation is occuring from.

With these two remarks, it should not be surprising that the statement of Proposition 6 holds. Of course it should be noted that the (common) discount factor does have a role in determining the consumption guarantee levels. As a result, even though the formulas for computing the equilibrium remain the same as those in Proposition 1, the eventual numbers will indeed be different. The larger the lapsation probability δ, the closer the equilibrium contract will be to the static equilibrium. ■

Having analyzed the consequences of lapsation into uninsurance for the shape of the equilibrium contracts, we close this appendix with a discussion of the ways in which such lapsation impacts welfare. There are two channels through which welfare is impacted:

First, lapsation into uninsurance will directly impact welfare by depriving the consumer of insurance at histories when such lapsation takes place (and onward). We do not quantify this direct effect here as one should expect it under all possible insurance regimes, not just long-term contracts.

The second channel through which lapsation into uninsurance changes welfare is through its impact on the shape of the equilibrium contracts as theoretically characterized by Proposition 6. We do not empirically quantify this channel either but we emphasize that a lower bound for the welfare from long-term contracts with a γ chance of lapsation into uninsurance would be the welfare resulting from myopia with $\beta = \delta \times (1-\gamma)$. This is the case because in the case of such myopia it is only the consumer's discount factor that decreases by a factor of $1-\gamma$; whereas in the case of exogenous lapsation, the insurer's discount factor decreases as well, which makes future protection cheaper for the consumer by requiring less frontloading. Given this lower bound and given that our analysis in the main text suggests long-term contracts provide non-trivial protection against reclassification risk even under severe myopia, we expect them to also have non-trivial performance under lapsation into uninsurance.

We close this appendix by noting that in the above analysis, the fact that the exogenous lapsation was into uninsurance does not have a critical role. Proposition 6 would also apply if exogenous lapsations were into employer-sponsored insurance instead of uninsurance.

9 Appendix I: Income Uncertainty

Our analysis in the main text does not require insurers to know the income profile of each consumer. Our self-selection analysis shows that for the main result to hold, it would suffice that the consumer know about his income profile. We did not, however, discuss the case where the consumer himself faces uncertainty about his income profile. This appendix provides a discussion of the issue. More specifically, we consider two cases. First, we examine cases where income realizations, like health, can be observed by both the insurer and the consumer and can be contracted upon. In other words, this would be a case where health insurers can also provide a form of income insurance. Second, we briefly discuss the case where income-uncertainty exists but income contingencies cannot be contracted upon as part of a long-term health insurance plan.

9.1 Income uncertainty under contractible income

If income is contractible, then risk in income can be treated in the same way as risk in health-care expenses. In other words, most of our characterizations of the equilibrium contracts will survive an extension of the model to income uncertainty. This includes the main result introducing consumption guarantees, the computation method, and the extensions and myopia and inertia. This does not, however, include the self-selection result as we are now assuming income is contractible. It is worth noting a key feature of our characterizations which makes them robust to inclusion of income uncertainty. That feature is the flexibility in the evolution of uncertainty over time. As mentioned in the proofs, we do not impose any restriction on how health status stochastically evolves over time. If we had any strong restriction (e.g., if it had to be a 1st order Markov process) then it would not be possible to readily extend the results to income uncertainty.

In terms of practical implementability, a combined health-and-income insurance program may seem far fetched. That said, we believe "coarse" versions of such dual insurance are not unimaginable. For instance, the consumer may frontload some amount of money in the initial periods and, in exchange for that, purchase future protection against increased premiums due to large health shocks under one condition: he will be entitled to the purchased protection in those future periods *only if he will be able to provide proof of belonging to a certain low-income bracket.*

9.2 Income uncertainty under non-contractible income

If income uncertainties on the part of the consumer exist but their realizations are not contractible with the insurer, then our characterization of equilibrium contracts may no longer hold. We leave developing such a characterization to future research and confine to an informal discussion of our conjectured implications of our current characterization for environments with uncertain but non-contractible income.

We expect contracts of the consumption-guarantees form we have characterized to provide non-trivial protection against reclassification risk even in environments with non-contractible income uncertainty (where they are no longer theoretically optimal). Suppose the insurer is offering a set of long-term contracts, each optimal for a certain income profile if the consumer is certain he will have that exact profile. Intuitively, a consumer uncertain about his future income may still benefit from signing the optimal contract for one of the possible future income profiles he can have. Which of the offered contracts (or, equivalently, "representative" income-profile) he will go with would depend on the likelihood of each possible income profile for that specific consumer as well as the amount of protection he would need under each such profile.

As one instance, a consumer who is uncertain whether he will have a flat income or an income path that will increase \$2K/year may choose to purchase a long-term contract that is optimal for a consumer who is certain his income will grow by \$$x$K per year where $x \in (0, 2)$. We would expect x to be closer to 0 than to 2 because the flat-income future self of this consumer will, by the concavity of the utility function, get a higher weight in the consumer's decision making.

Though we do not provide any quantitative implementation of the above idea, it is worth noting that the analysis in Atal et al (2020) is in line with our reasoning here. They examine a setting in which only the optimal contracts for flat-income consumers are offered; while in their setting, there are two types of income paths both having a non-trivial increase over time. They show that the flat-income based contracts do well to protect those consumers against reclassification risk.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Income profile	C^*	C^*_{NBNS}	CE_{SPOT}	CE_D	$\frac{C^*_{NBNS}-CE_{SPOT}}{C^*_{NBNS}}$	$\frac{CE_D-CE_{SPOT}}{C^*_{NBNS}-CE_{SPOT}}$	$\frac{CE_D-CE_{SPOT}}{C^*-CE_{SPOT}}$
Flat net	54.22	54.22	44.14	50.38	0.186	0.619	0.619
Non-mngr	54.22	47.33	37.10	38.76	0.220	0.162	0.097
Manager	84.73	55.65	45.49	46.16	0.183	0.066	0.017
Downs Mngr	54.22	37.65	27.45	28.59	0.271	0.112	0.043

Table 20: Replacating our baseline welfare results using a first-order Markov model for estimating transition probabilities.

10 Appendix J: The Role of Health Status Transition Matrix

In this appendix, we study how important it is to accurately estimate the health status transitions. In particular, we ask whether we would get similar welfare results if we approximated the health status transitions using a first-order Markov process instead of a second-order one. To this end, we produce an equivalent of our baseline welfare table (Table 7) but this time using first-order Markov estimates for the transition matrices. That is, we estimate new transition probabilities that assume the probability distribution over health status λ_{t+1} should only be conditioned on the realized value for λ_t, rather than on both λ_t and λ_{t-1}. This limits the ability of the model to capture the extent to which health status persists over time. We then use these estimates to simulate new equilibria and assess welfare. These welfare results are shown in Table 20.

As can be seen by examining columns 6 and 7 from this table, the welfare results under a first-order Markov assumption for transition probabilities are, in the case of all income paths, higher than their counterparts from our baseline results in Table 7 by about 50%.[15] This comparison shows that it is indeed important to estimate the health status transition process more accurately.

The direction of the comparison between the two sets of welfare results is also worth analyzing. We argue that the reason behind this comparison is that our estimated second-order Markov process exhibit more persistence in the health status relative to the first-order Markov estimates.[16] The equivalent of the bottom panel of Table 3 under first-order Markov transitions is shown in Table 21 verifies that these new transitions indeed lead to substantially

[15]Note that columns 1, 2, 3, and 5 are also slightly different between the two tables. That there is some difference here should not be surprising given that transition matrices impact these columns as well. But these differences are not economically meaningful and are substantially smaller than the difference between the two tables in columns 4, 6, and 7.

[16]Note that both first- and second-order transitions are capable of capturing full persistence (i.e., $\lambda_{t+1} = \lambda_t$ with probability 1) and no persistence at all (i.e., λ_{t+1} and λ_t being independent random variables). As

Health status	Ages			
(age 29 and 30)	30	31-35	36-40	41-64
1	0.84	2.43	3.08	4.64
4	3.05	3.5	3.18	4.65
7	20.51	6.68	3.32	4.65

Table 21: This table is similar to the bottom panel of Table 3 except that it is based on the less persistent, first-order Markov transitions. It reports, for various age ranges, the constant annual medical expenses (in thousands of dollars) such that the present discounted value of these constant annual expenses equals the expected present discounted value of expenses over the age range in question for a Utah man in various age-30 health states.

less persistent health states. Under more persistence, insuring against reclassfication risk using long-term contracts is more difficult. This is because, intuitively, long-term contracts work by allowing young and healthy individuals to "send money to their future sick selves" by frontloading. When there is more persistence, those who need to do so are more likely to be currently sick, lowering their incentive to frontload.

a result, we are not making a general claim that second-order Markov transitions are more persistent. Nevertheless, they seem to be in our context.